Journal of Experimental Psychology

JAN 18 1955

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

SOCIAL SCIENCE LIBRARY

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Retroaction as a Function of Discrimination and Motor Variables: M. L. RITCHIE AND F. A. MUCKLER 409

The Effect of Delay on Simultaneous and Successive Discrimination in Children: M. J. PERKINS, H. P. BANKS, AND A. D. CALVIN 416

Retention Functions in Reproductive Inhibition: G. E. BRIGGS, R. F. THOMPSON, AND W. J. BROGDEN 419

The Effect of Partial Reinforcement on Acquisition and Extinction of a Class of Verbal Responses: F. H. KANFER 424

Stimulus Generalization as a Function of Verbal Reinforcement Combinations: A. H. BUSS, M. WEINER, AND E. BUSS 433

Dimensional Analysis of Psychomotor Abilities: E. A. FLEISHMAN 437

Some Effects of Work Loading in a Repetitive Motor Task: I. McD. BILODEAU AND E. A. BILODEAU 455

Tilt Adaptation and Figural After-Effects: E. G. HEINEMANN AND T. MARILL 468

The Effect of Training on Absolute Estimation of Distance over the Ground: E. J. GIBSON AND R. BERGMAN 473

S-R Compatibility: Correspondence among Paired Elements within Stimulus and Response Codes: P. M. FITTS AND R. L. DEININGER 483

Autokinetic Movement of Very Large Stimuli: W. EDWARDS 493

The Reconstruction of Abbreviated Printed Messages: A. CHAPANIS 496

American Psychological Association

Vol. 48 No. 6 December 1954

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1954 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright, 1955, by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 48, NO. 6 DECEMBER, 1954

RETROACTION AS A FUNCTION OF DISCRIMINATION AND MOTOR VARIABLES[1]

M. L. RITCHIE AND F. A. MUCKLER

University of Illinois

Military training can benefit from psychomotor experiments when the experimental results may be applied to real training. The major problem which arises in making such application is that the results from different experimental apparatuses frequently lead to contradictory conclusions. When this situation occurs it is said that the results are functions of the particular apparatus. A logical next step for research is to discover what characteristics of a particular apparatus are responsible for the training results obtained.

Interpreting experimental results depends upon some kind of conceptual system. In this experiment an attempt will be made to evaluate the utility of a conceptual scheme[2] in which a discrimination and a motor stage are defined in a given motor task. The discrimination stage is defined as the election of the particular movement to be made. The motor stage is defined as the execution of the movement. If making changes in one stage produces different results from changes in the other stage, we may hope to relate certain apparatus characteristics to each stage.

An apparatus has been developed to provide a task in which separate changes may be made in the discrimination and motor stages. Retroaction results will be presented for each stage as functions of the amount of original learning and amount of interpolated learning.

Following the terminology of Osgood (**5**, pp. 520–522), retroaction is operationally defined as a comparison of performance on the last part of original learning with the first part of relearning in the paradigm OL-IL-RL. If the RL scores represent an improvement over OL the change is called retroactive facilitation. If the change is a decrement the term retroactive interference is used.

METHOD

Apparatus and experimental task.—The *S* was presented with a vertical stimulus panel containing two square arrays of four colored lights each. Seated approximately 6 ft. from the

[1] This research was supported in part by the United States Air Force under Contract AF 33(038)-25726, monitored by the Air Force Personnel and Training Research Center. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

[2] We are indebted to Professor C. E. Osgood for this basic idea and many helpful suggestions.

stimulus panel, *S* operated a response unit of four two-way switches. The discrimination task was the perception of the *order* in which the four different colored lights (red, green, blue, and white) were presented. The order of presentation was varied from trial to trial in the 24 possible combinations to eliminate as far as possible the anticipation of sequence. Finally, the motor task involved either pushing down or lifting up (and then pushing down) one of the four two-way switches to turn on a particular colored light.

The apparatus was specifically designed to provide a means of varying separately, in interpolated learning, the discrimination task and the motor task in the same apparatus. Accordingly, for the *discrimination interpolated learning task*, the discrimination task was altered by changing the contacts on the response switches so that a given switch turned on a different colored light. Specifically, during the original learning, the switches, from left to right, actuated the red, green, white, and blue lights. For interpolated learning, the two pairs were changed; that is, the switches, from left to right, actuated the green, red, blue, and white lights.

During original learning under all conditions, *S* was required to press down on a switch to actuate each stimulus light. This response was the same throughout the discrimination task. During the *motor interpolated task S* normally held all the switches down. A light was actuated by allowing the switch to rise and then depressing it again.

Finally, the relearning trials were simply a repetition of the original learning conditions.

Experimental design.—The *S*s were assigned at random to the experimental groups. All the experimental conditions are summarized in Table 1 with the primary error data. The following relationships were effected within the 18 groups. Groups 1–8 represent variations in original learning trials with the number of interpolated learning trials held constant at 48. In addition, alternate groups represent variations in the type of interpolated task, one being a discrimination task and the other a motor task. Groups 9–18 represent variations in interpolated learning trials with the discrimination and motor tasks and with original learning held constant at 72 trials. Groups 11 and 12 serve for both variations in original learning and inter-

TABLE 1

MEANS AND *SD*'S OF ERROR SCORES

Group	Total Trials		OL Trials				IL Trials		RL Trials*			
	OL	IL	13–24		Last 12		1–12		1–12		12–24	
			Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1D†	24	48	6.6	2.46	6.6	2.46	8.7	3.59	7.1	2.11	1.8	1.74
2M‡	24	48	6.6	3.72	6.6	3.72	5.8	3.68	3.0	2.93	2.8	2.27
3D	48	48	6.3	3.74	4.0	3.26	7.8	3.25	6.2	4.02	3.6	2.90
4M	48	48	5.6	3.53	3.0	2.93	4.3	1.27	0.9	1.04	1.3	1.55
5D	96	48	5.8	2.60	1.8	1.89	6.0	2.57	4.1	1.81	2.2	1.83
6M	96	48	9.2	3.43	4.2	2.09	5.4	2.33	2.9	2.21	2.5	1.75
7D	144	48	5.1	3.24	1.7	1.85	7.9	1.76	5.1	2.51	3.1	2.17
8M	144	48	5.7	3.46	1.4	1.62	2.9	2.12	1.6	1.43	1.2	1.17
9D	72	24	7.6	3.37	2.8	3.25	7.1	3.05	4.0	2.49	2.0	2.47
10M	72	24	4.9	3.21	1.7	1.05	3.5	1.75	2.0	1.27	0.7	0.64
11D	72	48	5.2	3.06	1.5	2.72	7.2	2.52	4.2	1.83	1.5	1.12
12M	72	48	5.4	2.38	1.1	1.14	3.8	2.86	1.0	0.78	0.7	0.78
13D	72	72	7.0	2.32	3.0	2.61	8.6	2.54	5.3	3.13	4.5	3.17
14M	72	72	7.3	4.01	3.6	2.06	4.2	3.03	2.0	1.79	1.4	1.11
15D	72	96	5.2	3.57	1.0	1.10	6.7	3.13	5.0	2.90	2.9	2.95
16M	72	96	6.1	4.06	2.7	2.28	2.7	2.00	1.3	1.00	0.9	1.14
17D	72	120	5.9	4.06	2.0	1.61	7.6	2.01	5.5	1.86	4.6	3.20
18M	72	120	6.2	3.31	2.9	2.91	3.3	1.85	1.3	1.10	1.3	1.42

* Relearning was for 24 trials in all groups.
† Discrimination interpolated task.
‡ Motor interpolated task.

polated learning. For all experimental groups, relearning consisted of 24 trials under the conditions of original learning.

Procedure.—Each *S* was given the following instructions for original learning: "On the panel before you there are two separate groups of four colored lights. The four lights on your left will be turned on in some particular color order such as red, green, blue, and white. Your task is to turn on the group of four lights on your right in this same order, using the four switches on the board before you. The order of presentation will be different each trial. You will always be matching the order by color. Both time and errors will be scored. Be as accurate and as fast as you can be. Your four lights must be in the correct order for the trial to count as correct. Are there any questions?"

For the discrimination interpolated groups, the following instructions were given: "I have made some changes in the apparatus. You are to do the same task as in the first series. However, you will find that the switches now turn on different colors. It will be necessary for you to learn which switches now control which lights. Are there any questions?"

For the motor interpolated groups, the following instructions were given: "I have made some changes in the apparatus. You are to do the same task as in the first series. However, instead of pushing down on the response switch to turn on the light, I want you to hold down all four switches and lift up the switch to turn on the light. Put it down again after you have made your response. The switch turns on the same light as it did before. Are there any questions?"

For the relearning trials, the following instructions were given: "We will now return to that task which you learned in the beginning. Are there any questions?"

Each *S* performed the total task in one experimental session. A 1-min. time interval was introduced between OL and IL and IL and RL in order to make necessary apparatus changes. It was possible to complete the learning series in a maximum of 1 hr. per *S*. Under these conditions it is always possible that work and/or fatigue factors may enter as complicating variables although we found no evidence, either from *S*s' reports or observable decrements in performance, that indicated the operation of these factors.

Measures.—The *S* received four stimulus lights and made four responses. If the order of responses was incorrect, the trial was counted as an error. If the trial was correct, it was counted as correct, and, in addition, it was possible to record the time of the correct response.

Retroaction effect was measured by a comparison of the final original learning performance and the relearning or test performance. Each *S* served as his own control (5). The formula used was:

$$\% \text{ Retroaction} = \frac{\text{Final OL score} - \text{RL score}}{\text{Final OL score}} \times 100$$

Thus, the basis for computation of retroaction effect was the performance reached by the end of original learning.

Results

Two response measures were recorded: (*a*) error scores, and (*b*) the time score for each correct response. Arbitrarily, it was decided to analyze the responses in units of 12 trials. Calculations were based on the number of errors and the mean time score for each block of 12 trials. Selected means and *SD*'s for error scores are given in Table 1.[3]

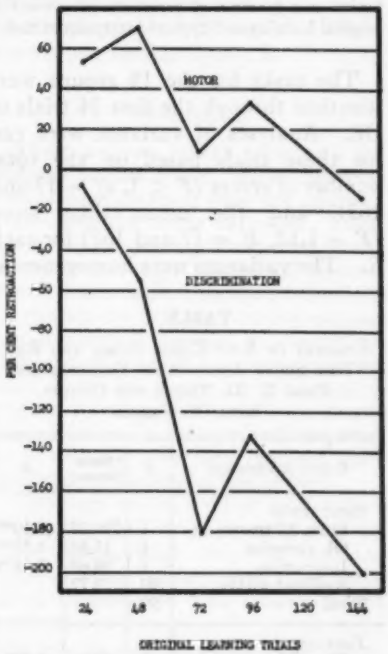


Fig. 1. Error scores: percentage of retroaction as a function of variations in amount of original learning and type of interpolated task

[3] Four tables giving data for (*a*) mean error scores for all experimental groups under all experimental conditions, (*b*) mean time scores in seconds for correct responses for all groups and conditions, (*c*) the computation of retroaction effect, and (*d*) selected means and standard deviations for time scores have been

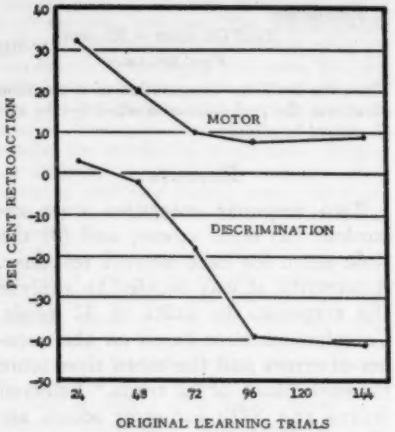


FIG. 2. Time scores: percentage of retroaction as a function of variations in amount of original learning and type of interpolated task

The tasks for the 18 groups were identical through the first 24 trials of OL. Analyses of variance were run for these trials based on the total number of errors ($F < 1$, $df = 17$ and 162), and the mean time score ($F = 1.12$, $df = 17$ and 162) for each S. The variances were homogeneous.

TABLE 2

SUMMARY OF RAW ERROR SCORE AND RAW TIME SCORE ANALYSIS OF VARIANCE ON FIRST 12 RL TRIALS FOR GROUPS WITH OL VARIED

Source of Variation	*df*	Mean Square	*F*
Error scores			
D vs. M treatment	1	298.29	80.19**
OL variation	4	15.88	4.27**
Interaction	4	29.69	7.98**
Residual within	90	3.72	
Total	99		
Time scores			
D vs. M treatment	1	10.96	6.52*
OL variation	4	5.05	3.01*
Interaction	4	1.04	—
Residual within	90	1.68	
Total	99		

* $p = .05$.
** $p = .01$.

deposited with the American Documentation Institute. Order Document No. 4412, remitting $1.25 for microfilm or $1.25 for photocopies.

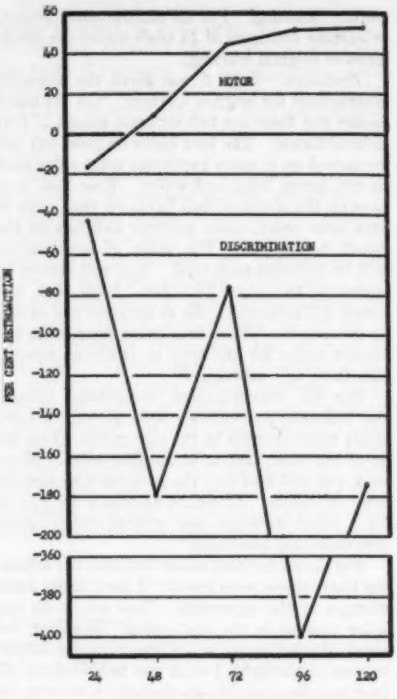


FIG. 3. Error scores: percentage of retroaction as a function of variations in amount of interpolated learning and type of interpolated task

We may conclude that the groups were drawn from the same population.

Variations in amount of original learning.—In Fig. 1, retroaction effect is shown as a function of variations in amount of original learning and type of interpolated task for error scores. In general, the functions appear to be (*a*) decreasing facilitation for the motor task, and (*b*) increasing interference for the discrimination task. It may be noted that the functions are similar in shape but that they fall on opposite sides of the zero point.

In Fig. 2, retroaction effect for time scores is shown as a function of variations in amount of original learning and the type of interpolated task. Again, the functions appear to be (*a*) decreasing facilitation for the motor

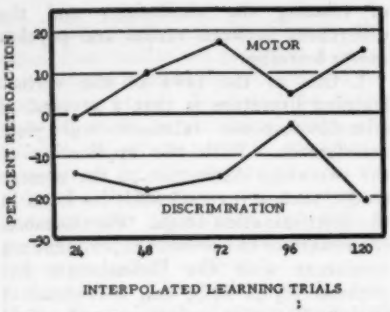


FIG. 4. Time scores: percentage of retroaction as a function of variations in amount of interpolated learning and type of interpolated task

interpolated task, and (*b*) slight facilitation to increasing interference for the discrimination interpolated task as amount of original learning increases. It may be noted that asymptotic levels appeared to have been reached for both functions.

Analyses of variance were performed on the raw scores for the first 12 trials of RL. These analyses are summarized in Table 2.

Variations in amount of interpolated learning.—Figure 3 shows the percentage of retroaction for error scores as a function of variations in amount of interpolated learning and type of interpolated task. It should be noted that quite different functions were obtained. With the increase of degree of learning of the motor interpolated task, increasing facilitation occurred, reaching an asymptotic level. Increasing the degree of interpolated learning for the discrimination interpolated task shows increasing interference.

Figure 4 shows the retroaction effect for time scores as a function of variations in amount of interpolated learning and type of interpolated tasks. The motor IL effects are facilitative while the discrimination IL effects are interfering.

Analyses of variance were run on the raw scores of the first 12 RL trials for the groups with IL varied. These are summarized in Table 3.

TABLE 3

SUMMARY OF RAW ERROR SCORE AND RAW TIME SCORE ANALYSIS OF VARIANCE ON FIRST 12 RL TRIALS FOR GROUPS WITH IL VARIED

Source of Variation	*df*	Mean Square	*F*
Error score			
D vs. M treatment	1	268.96	62.26**
IL variation	4	3.13	—
Interaction	4	3.38	—
Residual within	90	4.32	
Total	99		
Time score			
D vs. M treatment	1	10.19	7.66**
IL variation	4	1.57	1.18
Interaction	4	1.66	1.25
Residual within	90	1.33	
Total	99		

** $p = .01$.

DISCUSSION

There are two salient aspects to the findings of this study. First, different results were obtained from the discrimination and the motor variables. These results may be readily interpreted by the use of a two-stage conceptual scheme. When this is done there are implications for the relation between psychomotor and verbal learning. Second, the discrimination and the motor functions obtained were of similar shape for variations in original learning, but of different shapes for variation in interpolated learning. It is believed that these findings have implications for retroaction theory.

Two-stage analysis.—The results show clearly that facilitation resulted from the motor interpolated task and that the discrimination interpolated task produced interference. An analysis which postulates two stages in the psychomotor task may be used to explain these results. The two stages involve (*a*) the selection of movements to be made and (*b*) the execution of these movements. The first stage is defined as the *discrimination stage* which in this experiment is the selection of the sequence in which the four fingers are to be moved. The second is the *motor stage* which is the execution of the movements once the

sequence of fingers to be moved has been selected.

In applying this terminology to the "star discrimeter" (1), for example, the discrimination stage consists in the selection of the slot into which the lever is to be moved. The motor stage consists in the actual movement of the lever into the slot. For the pursuit rotor the discrimination stage is simply electing to keep the stylus on the dot. The complex of interrelated movements required to keep the stylus in place comprises the motor stage. An analysis of the discrimination stage in this experiment will show that the probability that *S* will select the correct sequence by chance alone is 1 in 24. The task of the discrimination stage is defined for *S* as that of selecting the one correct sequence from 24 alternative sequences. When the discrimination interpolated task is introduced *S* now finds these 24 alternatives rearranged. An important factor in determining interference effects produced will be the number of alternatives involved in the discrimination stage.

A similar analysis of the motor stage shows that a probability situation like that of the discrimination stage is not present. The motor stage has been defined as the execution of the finger movements once the sequence of fingers to be moved has been selected. In original learning each finger rests on a key and *S* is instructed that execution involves depressing the key. In the motor interpolated task *S* holds the keys down and releases each to actuate the response. It is to be noted here that when *S* begins the interpolated task the motor response of original learning (depressing the key) is rendered a *physical impossibility* by the apparatus design. These operations are very similar to those of Wickens (6, 7) in his studies of response generalization where the use of a hand flexion response during transfer learning made the occurrence of the originally learned extension response impossible.

Psychomotor vs. verbal learning.—It has been frequently said that generalizations from verbal learning may not be applied to psychomotor tasks. It appears that a two-stage analysis of the present experiment provides us with a concrete means of relating the similarities and the differences between verbal and psychomotor learning.

1. One of the laws of the verbal-learning literature is that a reversal of stimulus-response relations will yield interference. With the application of the two-stage distinction to the present study, such a reversal finds its locus in the discrimination stage. Furthermore, the results of the present experiment are consistent with the Poffenberger hypothesis (8, p. 202), i.e., a reversal of stimulus-response relations will yield interference.

2. The limitations imposed by the apparatus may be involved also in accounting for a difference between the predictions of verbal-learning laws and psychomotor results.

It will be observed that the stimulus-response paradigm for learning a new response to an old stimulus is descriptive of the motor interpolation variable in the present study. The verbal-learning results from this paradigm have shown interference effects (8, p. 202), while the results from the present study show facilitation. In the analysis of the motor stage of the present study given above, it was observed that the apparatus itself prevents *S*, in interpolated learning, from making the original learning response. Thus, it would be expected that interference would not result from this change in the motor task, and that the facilitation obtained is consistent with that shown by Wickens (6, 7).

When the paradigm for learning a new response to an old stimulus has been used in verbal learning it has usually involved changes in the discrimination stage and not in the motor stage. That is, the substitution has usually involved learning to select a new word and not learning a new motor expression of a word whose selection habit has already been formed. It is possible that verbal studies of meaningful similarity have involved changes in more than one stage.

3. One of the important differences suggested in the present analysis has to do with the probability of possible alternatives. In this experiment it was noted that *S* was required to choose one response sequence from 24 possible

alternative sequences. In this case, the retroaction results are probably a function of the number of alternatives and the probability state. It should be noted that in most psychomotor tasks the number of alternatives is specifically determined by the apparatus and the task. In the case of verbal learning, however, the probability function is generally unknown since the alternatives rest in a language matrix.

Retroaction theory.—Lewis (3) concludes from a series of experiments that both facilitation and interference effects are involved in the effects of interpolated activity in psychomotor learning. Although cautious about direct comparisons with verbal learning, Lewis sees his results as not inconsistent with the verbal-learning experiments which led to a two-factor explanation (4) of retroactive inhibition. Both Lewis' work and the verbal-learning literature show monotonic increasing functions of interference with the amount of original learning. However, for increases in interpolated learning, Lewis was able to show only monotonic increasing interference while the verbal-learning studies show an increase in interference followed by a decrease. Lewis concludes that a two-factor theory for psychomotor learning might be required if the decreasing trend also is discovered in a psychomotor task.

The results of the present experiment are consistent both with Lewis' results and with verbal learning for variations in amount of original learning. For increasing amounts of interpolated learning, however, very different results were obtained. The motor interpolated task resulted in increasing facilitation while the discrimination interpolated task resulted in increasing interference. Thus, two separate and opposite trends result from an increase in interpolated learning.

If subsequent experimentation confirms the present *tentative* findings, then it appears that Lewis is correct in his supposition that two factors will be needed to account for them. Furthermore, it appears not unreasonable to hope that the factors involved may be tied down quite closely to apparatus and task characteristics.

Summary

In a retroaction experiment, three variables were investigated: (*a*) variations in the nature of the interpolated tasks, (*b*) variations in amount of original learning, and (*c*) variations in amount of interpolated learning.

The interpolation of a discrimination task resulted in interference while interpolation of the motor task gave facilitation. Increasing the amount of original learning resulted in decreasing facilitation for the motor interpolated task and increasing interference for the discrimination interpolated task. Increasing the amount of interpolated learning resulted in increasing facilitation for the motor interpolated task and increasing interference for the discrimination interpolated task.

The implications of these results are discussed for (*a*) a two-stage analysis of psychomotor performance, (*b*) differences between verbal learning and psychomotor learning, and (*c*) retroaction theory.

References

1. Duncan, C. P. Transfer in motor learning as a function of degree of first-task learning and inter-task similarity. *J. exp. Psychol.*, 1953, **45**, 1–12.
2. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
3. Lewis, D., & McAllister, D. E. Facilitation and interference in performance on the modified Mashburn apparatus: II. The effects of varying the amount of interpolated learning. Port Washington, L. I., N. Y.: U. S. Navy, Special Devices Center, 1950. (Tech. Rep.—SDC 938-1-2.)
4. Melton, A. W., & Irwin, J. McQ. The influence of the degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 175–203.
5. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univ. Press, 1953.
6. Wickens, D. D. Studies of response generalization in conditioning. I. Stimulus generalization during response generalization. *J. exp. Psychol.*, 1943, **33**, 221–227.
7. Wickens, D. D. Studies of response generalization in conditioning. II. The comparative strength of the transferred and non-transferred responses. *J. exp. Psychol.*, 1943, **33**, 330–332.
8. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received March 29, 1954)

Journal of Experimental Psychology
Vol. 48, No. 6, 1954

THE EFFECT OF DELAY ON SIMULTANEOUS AND SUCCESSIVE DISCRIMINATION IN CHILDREN[1]

M. J. PERKINS, H. P. BANKS, AND ALLEN D. CALVIN

Michigan State College

A large number of articles have recently appeared regarding the relative difficulty of simultaneous and successive discrimination. A simultaneous discrimination is fundamentally the standard discrimination problem where both the positive and negative cues are present on each trial, and *S* must learn to choose the correct cue regardless of spatial arrangement. The successive discrimination is one where only a single cue is present during a trial, and *S* must make a spatial response on the basis of this cue.

Although there has been a great deal of work done in this area, only one study relating to this problem has appeared in the literature where humans were used as subjects. Loess and Duncan (1) used college students as *S*s and reported that there was no significant difference between successive and simultaneous discrimination on an easy problem, but on a more difficult one the simultaneous discrimination was easier to master.

The question of the importance of delay in simultaneous and successive discrimination is suggested by a paper presented by Mishkin and Pribram (2) at the 1953 APA convention. They found that monkeys with bilateral frontal lesions learned several types of discrimination problems in a delay situation that required a response to a nonspatial cue, but failed to solve a problem that called for a traditional spatial response. Monkeys with bilateral temporal lesions, which were used as controls, solved all the problems presented.

The present experiment was designed (*a*) to study the effect of simultaneous and successive presentation of stimuli in children and (*b*) to examine the interaction between delay and method of stimulus presentation.

METHOD

Subjects.—The *S*s were 56 children from the third and fourth grade of a local public school. They were randomly divided into four groups of 14 each, except for the restriction that each group consist of 8 *S*s from the third grade and 6 from the fourth.

Apparatus.—The apparatus was a plywood stage with a cardboard panel which could be raised and lowered by *E*, two identical cups under which small plastic toys were placed to reward *S* when a correct choice was made, and two stimulus cards, one white with five stars—one in each corner and one in the center—and the other plain black. These cards were selected on the basis of an earlier experiment.

Treatments.—Group I, the simultaneous nondelay group, was presented with a standard discrimination problem with both the positive and negative cues present on each trial. Half of the *S*s had black positive and half had white positive, and the stimuli were presented in a modified random order such that no stimulus card ever appeared in the same position three times in succession. The procedure for Group II, the simultaneous delay group, replicated that for Group I except that *S*s in this group were allowed to look at the stimuli for 5 sec. and then the cardboard panel was dropped for 5 sec. during which the stimuli were removed. After this 5-sec. delay period the panel was raised and *S*s made their choice. Group III, the successive nondelay group, was presented with only a single stimulus card on each trial

[1] The authors wish to express their thanks to C. E. MacDonald, Superintendent of Schools, East Lansing, Michigan, and Noel M. Ranger, Principal of the Red Cedar School in East Lansing, for their kind cooperation which made this study possible.

and had to make a spatial response on the basis of that cue, i.e., when the black card was presented *S* had to learn to choose the left cup and vice versa. The same type of balanced experimental design described previously was utilized. The procedure for Group IV, the successive delay group, was the same as that of Group II except that the stimuli were presented successively as set forth above.

Each *S* was taken individually into the testing room and given the following instructions: "We are going to play a game. I am going to hide a toy under one of these cups. If you guess which cup it's under, you get to keep the toy. Would you like to ask any questions?" The criterion for solution was set at ten successive correct choices. If *S* had not solved the problem after 40 presentations, the trials were terminated.

Results and Discussion

Since the trials were terminated if the *S* had not solved after 40 presentations, we have presented the results in terms of median trials to learn in Table 1.

TABLE 1

Median Trials to Learn

Group	Trials	No. of *Ss* Who Failed to Solve
I. Simultaneous nondelay	12.5	0
II. Simultaneous delay	15.5	2
III. Successive nondelay	37.5	7
IV. Successive delay	40.0	8

In order to determine the effect of simultaneous and successive presentation, we made the following comparisons utilizing Wilcoxon's nonparametric method (7, p. 4): Group I (simultaneous nondelay) vs. Group III (successive nondelay), Group II (simultaneous delay) vs. Group IV (successive delay), and I and II combined vs. III and IV combined. The difference between Group I and Group III was significant beyond the 5% level. The difference between Group II and Group IV was significant beyond the 2% level. The difference between Groups I and II combined and Groups III and IV combined was significant beyond the 1% level. Thus the simultaneous discrimination was significantly easier under both conditions lending empirical support to the theoretical position taken by Spence (6) and Nissen (3, 4, 5). A χ^2 test for interaction between delay and method of stimulus presentation fell far short of significance, and tests of significance (7, p. 4) between Groups I and II, III and IV, and Groups I and III combined vs. II and IV combined revealed no significant differences. These findings indicate that the effect of delay did not operate differentially in the simultaneous and successive problems, and that although delay tended to make both problems more difficult, the differences failed to reach significance. (The restriction in range for Groups III and IV of course makes their comparison quite tenuous.)

A test for *E* differences also was insignificant.

Summary

Fifty-six third and fourth grade children were divided into four groups of 14 each. Group I was given a simultaneous discrimination. Group II was given a simultaneous discrimination under delay conditions. Group III was given a successive discrimination. Group IV was given a successive discrimination under delay conditions. The principal findings were: (*a*) the simultaneous discrimination was easier under both conditions; (*b*) although delay did make both discriminations more difficult, the difference between the delay and nondelay groups was not significant, and the effect of delay did not operate differentially in the simultaneous and successive problems.

References

1. Loess, H. B., & Duncan, C. P. Human discrimination learning with simultaneous and successive presentation of stimuli. *J. exp. Psychol.*, 1952, **44**, 215–221.
2. Mishkin, M., & Pribram, K. H. Analysis of the effects of frontal lobe damage in

monkeys: I. Variations of delayed response. *Amer. Psychologist*, 1953, **8**, 405. (Abstract)

3. NISSEN, H. W. Description of the learned response in discrimination behavior. *Psychol. Rev.*, 1950, **57**, 121–131.
4. NISSEN, H. W. Further comment on approach-avoidance as categories of response. *Psychol. Rev.*, 1952, **59**, 161–167.
5. NISSEN, H. W. Approach and avoidance: A reply. *Psychol. Rev.*, 1952, **59**, 237–238.
6. SPENCE, K. W. The nature of response in discrimination learning. *Psychol. Rev.*, 1952, **59**, 89–93.
7. WILCOXON, F. *Some rapid approximate statistical procedures*. Stamford, Conn.: American Cyanamid Co., 1949.

(Received March 15, 1954)

Journal of Experimental Psychology
Vol. 48, No. 6, 1954

RETENTION FUNCTIONS IN REPRODUCTIVE INHIBITION[1]

GEORGE E. BRIGGS,[2] RICHARD F. THOMPSON, AND W. J. BROGDEN

University of Wisconsin

Underwood (12) and Briggs (1) have investigated reproductive inhibition as a function of time between interpolated learning (IL) and the measurement of retention. The learning materials were paired two-syllable adjectives. A single list of stimulus adjectives was used for both original learning (OL) and for IL, but the response adjectives were different for OL and IL. Retention of the response words for both OL and IL was measured by a technique introduced by Underwood (12) and called modified free recall (MFR). After OL and IL, *S* is presented with the stimulus words following instructions to respond with whichever of the two response words is first recalled. This technique provides for measurement of the relative strength of the associations formed in OL and IL.

The results of these two studies are different. Underwood found a non-significant increase in the percentage of response words from OL between times of 1 min. and 5 hr. following IL, whereas Briggs found a significant increase in such responses from 4 min. to 6 hr. Both investigators found a significant decrease in percentage of response words from IL between the comparable time intervals. For the longer time intervals, the results of the two studies are essentially in agreement.

The present study was designed to investigate the course of the two response systems during the first 6 hr. following IL. Its purpose was to determine the form of the functions obtained by Underwood and Briggs for short intervals not studied by them, and, if possible, to account for the discrepancy in their results at the 5–6 hr. interval.

[1] Supported in part by a grant from the Research Committee from funds provided by the Wisconsin Alumni Research Foundation.

[2] Now postdoctoral fellow of the National Science Foundation at Northwestern University.

Procedure

Six lists of ten common two-syllable adjectives were used for the stimulus words, and 12 lists of ten common two-syllable adjectives for the response words. A given list of stimulus words was used for both OL and IL, but there were different lists of response words for OL and IL. MFR was used only to measure retention and followed IL after 4, 12, 36, 108, 234, or 360 min. Each *S* was tested at each of these six intervals and was given a different list for each interval. The experimental design is that of a 6×6 latin square in which the letters represent the interval between IL and MFR, the rows the sequence of time intervals, and the columns the ordinal position of time intervals. Each time interval occurs once and only once in each row and column of the square. List was confounded with ordinal position.

The following treatment holds for each cell of the latin square. The *S* was instructed to free associate to the ten stimulus words presented serially. OL was initiated and terminated when the criterion of one perfect repetition was reached. IL then occurred and was terminated with the criterial trial. MFR followed IL at the appropriate interval. During both OL and IL, each stimulus word was presented alone for 4.5 sec. After .5 sec., the stimulus and response words were presented together for 4.5 sec. After .5 sec., the next stimulus word was presented, after an interval of .5 sec. this stimulus word and its response word were presented for 4.5 sec., and so on to the end of the list. The *S*s were required to respond during the 4.5-sec. stimulus period in order to be scored with an anticipation. Three orders of presentation of each list were used serially. The intertrial interval was 10 sec. The interval between free association and OL and that between OL and IL was 60 sec. During free association and MFR, each stimulus word was presented for 4.5 sec.

with an interval of .5 sec. between, and only one trial was given.

By utilization of a recording apparatus manufactured by the Tel-Autograph Corporation, two *Ss*, separated by a partition in the same room, learned the lists simultaneously and independently (1). The words were exposed separately for the two *Ss* on milk-glass screens by means of two La Belle automatic projectors. Each *S* wrote his response with the stylus of the recording apparatus. The 24 *Ss* were students in elementary psychology at the University of Wisconsin and had served previously in the Briggs study (1). The use of practiced *Ss* was intentional. Four *Ss* were assigned at random to each of the six sequences of the latin square, thus giving four replications.

At the beginning of the session, *S* was told that a list of ten adjectives would be presented, one at a time. He was instructed to write down the first word that came to mind when an adjective was presented on the screen. For OL and IL, *S* was instructed to write the response word during the period when the stimulus word was presented. He was also requested to draw a line through an incorrect response during the time the response word was presented. He was instructed not to write the response word during the time it was presented with the stimulus word. The *E* was able to check on this by watching the repeater stylus on the recording apparatus located in his room. The *S* was told the time relations between the stimulus and response words and notified that both OL and IL would terminate when the criterion of one perfect repetition of the list was reached. For MFR, *S* was told that the stimulus words would be presented alone, and was instructed to respond to each by writing whichever of the two responses assigned to the stimulus word was recalled first. If unable to recall either of those responses, he was told to write any word he thought of.

The records for MFR were read and the responses tabulated for each *S* for each time interval. The responses assigned to the first list are designated as R_1, those from the interpolated list R_2, those given during MFR from either list which were incorrect for the particular stimulus word as R_3, and, finally, those appearing in MFR which were from neither of the two lists, along with those appearing in free association, as R_0. The percentage of response in each of the above categories is the basic datum of the experiment. At no time during MFR did any *S* fail to respond with either an R_1, R_2, R_3, or R_0 written response.

Results

Table 1 gives the group mean frequency of response for the several response categories at each interval of retention. Since each *S* always made a response to the stimulus words in MFR, the sum of the frequencies of response for each category is 10 or 100%. Examination of the relations of percentage of R_0, R_1, R_2, and R_3 responses to retention time interval was made by means of an expanded analysis of variance technique for correlated data (8) and by the subsequent application of Duncan's range test (7) to the group means. The 5% level of confidence is the criterion of statistical significance. Because there was not homogeneity of variance of the raw data, the arc sin (per cent)$^{\frac{1}{2}}$ transformation was applied for purposes of analysis (11). The results of the analysis of variance on all categories of response are not presented, since application of Duncan's range test shows that neither the mean percentage of R_0 nor R_3 responses

TABLE 1

Group Mean Frequency of Response during MFR for the Several Categories at Each Interval of Retention

Response Category	Retention Interval											
	4 min.		12 min.		36 min.		108 min.		234 min.		360 min.	
	M	σ_M	*M*	σ_M	*M*	σ_M	*M*	σ_M	*M*	σ_M	*M*	σ_M
R_0	.17	—	.38	—	.33	—	.71	—	1.42	—	.96	—
R_1	3.46	.31	4.00	.36	3.79	.37	3.83	.30	3.79	.32	4.08	.24
R_2	6.13	.33	5.29	.43	5.58	.37	5.08	.32	3.83	.36	4.46	.33
R_3	.25	—	.33	—	.29	—	.38	—	.96	—	.50	—
$R_0 + R_3$	.42	—	.71	—	.63	—	1.08	.20	2.38	.39	1.46	.28

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE

Source	*df*	Mean Square	*F*
1. *a.* Sequence	5	68.02	.64
b. *Ss* within seq.	18	15.07	.14
2. *a.* Response categories	2	36,062.61	117.45*
b. S × RC	10	470.04	1.53
c. RC × *Ss* within seq.	36	307.04	2.91*
3. *a.* Retention intervals	5	55.37	.53
b. Ordinal position	5	20.58	.20
c. RI × RC	10	899.30	8.54*
d. OP × RC	10	282.22	2.68*
e. Square uniqueness	20	5.36	.05
f. SU × RC	40	140.29	1.33
g. OP × *Ss* within RC	270	105.36	—

* Significant at the .05 level of confidence.

differed significantly from each other as a function of retention interval. The frequency of such responses is small and the failure to obtain significant differences may be due to this. If the R_0 and R_3 categories are combined, this classification is the same as Underwood's "other" (12). Analysis of variance of the percentage of R_1, R_2, and $R_0 + R_3$ gives the same results for the R_1 and R_2 categories as the previous analysis, and, in addition, shows that the combination of $R_0 + R_3$ is affected significantly by retention interval. The summary of this latter analysis is given in Table 2. In addition to the usual sums of squares in the latin square analysis (9), the present design introduces Response Categories as another class of correlated data. Therefore, the analysis contains an additional term, Response Categories × *Ss* within Sequence, that may be used to test the mean square for response categories and the mean square for Sequences × Response Categories. This term may be used for such tests only if it is significant when tested against the basic error term, Ordinal Position × *Ss* within Response Categories. Since the Response Categories × *Ss* within Sequence mean square is significant when tested against the Ordinal Position × *Ss* within Response Categories mean square, it is used as the denominator of the *F* tests for the mean square for Response Categories and the mean square for Sequence × Response Categories. The mean square for Ordinal Position × *Ss* within Response Categories is the denominator for all other *F* ratios tested in Column 4 of Table 2.

The significant interaction of Retention Intervals × Response Categories indicates that the percentage of response in the response categories varies as a function of retention interval. Group means for each of these three categories of response are plotted for each retention interval in Fig. 1. More precise information about the form of these three functions and differences between them may be obtained from application of Duncan's range test to the 18 means. The significant differences that result from this test are given as follows. All the means in the R_1 and R_2 categories differ from all the means in the $R_0 + R_3$ category. The R_2 means for the 4-min. and 36-min. intervals differ from the R_2 means for 234 and 360 min. and from all the R_1 means. The R_2 mean for 12 min. differs from the R_2 mean for 234 min. and from all the R_1 means. The R_2 mean for 108 min. differs from all the R_1 means, except that for 360 min. There are

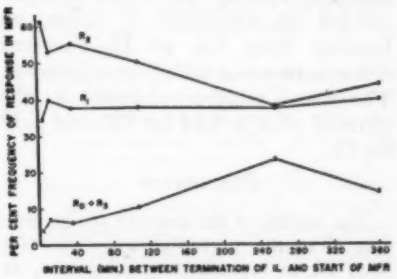


FIG. 1. Percentage of R_1, R_2, and $R_0 + R_3$ responses in relation to time interval between interpolated learning and modified free recall

no differences within the R_1 group of means. The $R_0 + R_3$ mean for 234 min. differs from those for 4, 12, 36, and 108 min., the $R_0 + R_3$ mean for 360 min. differs from those for 4, 12, and 36 min., and, finally, the $R_0 + R_3$ mean for 108 min. differs from those for 4 and 36 min. These results show, therefore, that the function for R_1 exhibits no change, the function for R_2 tends to decrease in a fairly regular manner, and the function for $R_0 + R_3$ tends to increase over the time intervals tested.

The other significant F ratios of Table 2 require comment. The ratio for response categories indicates that the mean percentage of response, over all retention intervals, differs significantly among categories R_1, R_2, and $R_0 + R_3$. The significant interaction of Ordinal Position × Response Categories probably represents differences in response category as a function of the different lists, since list is confounded with ordinal position in the experimental design, and the means vary nonsystematically with ordinal position.

In order to test for transfer from IL to OL, the difference in number of trials to attain the criterion ($N_{OL} - N_{IL}$) was obtained for each S for each retention interval. Because there was not homogeneity of variance, the data were transformed to arc sin $(N_{OL} - N_{IL} \times 7/13)^{\frac{1}{2}}$. The constant, 7/13, was added to eliminate negative values. Since the analysis yielded no significant F ratios, no transfer from OL to IL occurred either between or within time interval. The mean number of trials to the criterial trial is 5.34 for OL and 5.37 for IL.

Discussion

The results of the present study indicate that the percentage of responses from IL in reproductive inhibition, as measured by MFR, decreases significantly during the first 6 hr. after learning. The percentage of responses from OL shows no significant change during this period. The percentage of other responses, $R_0 + R_3$, increases significantly during the same interval. There is no significant transfer between OL and IL.

A comparison of the present study with those of Underwood (12) and Briggs (1) reveals the following. A significant increase in the percentage of responses from OL with time was obtained by Briggs, whereas in Underwood's study and the present study also there was a nonsignificant increase. Briggs found significant positive transfer from OL to IL. No significant transfer was found by Underwood or in the present study. All three studies, of course, found significant decreases in the percentage of responses from IL.

The three investigations differ in many respects. It should be noted, however, that in spite of the many differences between the present investigation and that of Underwood, the findings are essentially the same for both. If a comparison is made of the Briggs study, on the one hand, and the Underwood study and the present study on the other, there are four methodological differences that may account for the differences in results. (*a*) Briggs used a factorial design, whereas Underwood and the present investigators employed counterbalanced designs. (*b*) Briggs introduced MFR during OL or IL for his experimental group, while Underwood and the present investigators used it only after IL. (*c*) Briggs allowed a 24-hr. interval between OL and IL, whereas Underwood and the present investigators presented OL and IL in immediate succession. (*d*) Underwood and the present study used practiced *S*s, while Briggs did not.

The possibility exists that repeated measures on the same S influence MFR responses. However, the fact that no significant ordinal position effect was found in the present study would seem to indicate that this influence is neither significant nor systematic.

Concerning Briggs' introduction of MFR in OL or IL, it is conceivable that this amounts to the introduction of distributed practice. His experiment was

so designed that the effects of MFR could be tested. No such effect was found on relative retroactive inhibition as measured by the first *relearning* trial. Nevertheless, this does not completely exclude the possibility that the relative *strengths* of responses from OL and IL were changed. However, in view of the finding of Hovland (10) that paired-associate learning at a slow rate is quite insensitive to the effects of distribution, such an explanation would seem unlikely.

With regard to the increase in percentage of responses from OL, Briggs (1) suggests that this results from the positive transfer from OL to IL. Two possible sources of this transfer effect are the use of unpracticed *S*s and use of a 24-hr. interval between OL and IL. Concerning the former, it is quite likely that the unpracticed *S*s learned to a greater extent during OL than they did during IL, thus permitting an increase in OL responses as time from IL is increased.

The effect on transfer of introducing a 24-hr. interval between OL and IL is questionable. In a series of studies, Bunch and co-workers (2, 3, 4, 5, 6) investigated in some detail the course of transfer with time. With regard to the comparison of transfer immediately after learning to transfer after one or two days, they found either no change (2, 5) or a significant increase in positive transfer (3, 4, 6).

Thus it would seem likely that the differences between the Briggs study, on the one hand, and the Underwood study and the present investigation, on the other, can be attributed to Briggs' use of unpracticed *S*s and a 24-hr. interval between OL and IL. Direct experimentation would, of course, be required to verify this point.

Summary

The investigation was designed to determine the course of retention functions from original and interpolated learning as a function of relatively short time intervals between termination of interpolated learning and measurement of retention. The stimulus adjectives were common to original and interpolated learning, but the response adjectives were different. Retention was measured by means of the technique of modified free recall after intervals of 4, 12, 36, 108, 234, and 360 min.

Statistical analysis of the data in modified free recall shows that responses from original learning show no change, that responses from interpolated learning tend to decrease with time interval in a fairly regular manner, and that "other" responses tend to increase as the interval between termination of interpolated learning and modified free recall increases. These results are discussed in relation to relevant literature.

References

1. Briggs, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285–293.
2. Bunch, M. E. The amount of transfer in rational learning as a function of time. *J. comp. Psychol.*, 1936, **22**, 325–337.
3. Bunch, M. E. A comparison of retention and transfer of training from similar material after relatively long intervals of time. *J. comp. Psychol.*, 1941, **32**, 217–231.
4. Bunch, M. E., & Lang, E. S. The amount of transfer of training from partial learning after varying intervals of time. *J. comp. Psychol.*, 1939, **27**, 449–459.
5. Bunch, M. E., & McCraven, V. G. The temporal course of transfer in the learning of memory material. *J. comp. Psychol.*, 1938, **25**, 481–496.
6. Bunch, M. E., & Rogers, M. The relationship between transfer and the length of the interval separating the mastery of two problems. *J. comp. Psychol.*, 1936, **21**, 37–52.
7. Duncan, R. B. A significance test for differences between ranked treatments in an analysis of variance. *Virginia J. Sci.*, 1951, **2**, 171–189.
8. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
9. Grant, D. A. The latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, **45**, 427–442.
10. Hovland, C. I. Experimental studies in rote-learning theory. V. Comparison of distribution of practice in serial and paired associate learning. *J. exp. Psychol.*, 1939, **25**, 622–633.
11. Snedecor, G. W. *Statistical methods.* Ames, Iowa: Iowa State Coll. Press, 1946.
12. Underwood, B. J. 'Spontaneous recovery' of verbal associations. *J. exp. Psychol.*, 1948, **38**, 429–439.

(Received March 30, 1954)

Journal of Experimental Psychology
Vol. 48, No. 6, 1954

THE EFFECT OF PARTIAL REINFORCEMENT ON ACQUISITION AND EXTINCTION OF A CLASS OF VERBAL RESPONSES[1]

FREDERICK H. KANFER

Indiana University[2]

The effects of partial reinforcement on animal behavior have been explored extensively in numerous studies. These studies agree in their finding of greater resistance to extinction with partial groups than with continuous groups (9). Recently, some attempts have been made to extend these findings to human behavior. To date, however, the experimental studies with humans have been restricted either to classical conditioning (6) or the conditioning of "verbal expectations" (7). In the classical conditioning studies the procedure involves stimulus control by *E* and a reinforcement schedule independent of *S*'s responses. In the studies of "verbal expectations" the responses which *S* can make are similarly limited; *S* can make only two highly interdependent responses: either a "yes" response or a "no" response in anticipation of the presentation of a light stimulus.

In everyday human relationships behavior is usually modified by reinforcement which is contingent on the occurrence of a particular response class. Neither of the experimental procedures with humans, however, provides for such a contingency of reinforcement. The present study was designed to investigate the effects of partial reinforcement during acquisition and extinction of verbal behavior in a situation which allows for contingency of reinforcement on the occurrence of a specified class of responses. Also, a wide range of responses was used to minimize the interdependence of response classes. In these respects the present study is somewhat analogous to the situation which Skinner (11) describes for operant conditioning in animals and similar to everyday human behavior.

Methodological Considerations

An optimal demonstration of conditioning and extinction with operant conditioning techniques demands that the chosen response class have an initial probability of occurrence which is sufficiently high to expect a response from this class to occur within a reasonable time interval. It also must be low enough to allow for modifiability by the experimental operations. In addition, the response class must be clearly defined, i.e., it must delineate a pertinent aspect of *S*'s behavior. This requirement is especially difficult to meet when dealing with verbal behavior (3). One way to minimize these difficulties is to use numbers. While number behavior is subject to guessing habits (10) and previously learned patterning, the relationship between numbers is relatively simple and universally accepted. Despite individual differences, social training clearly establishes the order of magnitude of a class of number responses, defined by order of magnitude, and reduces the difficulties encountered with language behavior.

A situation which easily meets the criteria for use as an operant conditioning technique is the autokinetic situation. This situation permits a wide range of initial responses. The modifiability of responses in the autokinetic situation

[1] This article is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree in the Department of Psychology, Indiana University. The author is greatly indebted to Professors Irving J. Saltzman, Cletus J. Burke, and Bertram D. Cohen for their valuable aid during various phases of this study.

[2] Now at Washington University, St. Louis, Missouri.

has been shown repeatedly (4), and reinforcement can easily be made contingent upon preceding responses. The distribution of initial estimates of movement in the autokinetic situation shows large individual differences. It seems advisable, therefore, to use defining criteria of the class of reinforced responses which relate this class directly to each *S*'s distribution and which are not affected by individual differences or the numerical value of the estimates. This can be done by use of cumulative frequency distributions obtained from each *S*'s initial sample of responses. A class so defined can be experimentally manipulated until the frequency of its members reaches a desired value. Such a measure of responses reflects the probabilities of response evocation at the pre- and post-acquisition stage of the experiment. Furthermore it allows for more general conclusions than the frequently used method of adding a constant to the numerical mean of estimates of all *S*s (4).

The study of partial reinforcement presents a methodological problem which makes the interpretation of results difficult. Acquisition under conditions of partial reinforcement involves simultaneously at least four other variables: (*a*) the actual number of reinforcements given, (*b*) the total number of acquisitional trials allowed, (*c*) the number of "correct" responses that occur, regardless of subsequent reinforcement, and (*d*) the pattern of reinforcement employed. In most studies of partial reinforcement, therefore, the percentage of reinforcement is confounded with one or the other of these four variables. Consequently, the results of partial reinforcement studies frequently cannot be interpreted as due solely to the schedule of reinforcement. In the present study an attempt was made to isolate, by statistical means, the effects of the first three of the four variables mentioned above.

Method

Subjects and apparatus.—The *S*s were 124 volunteer undergraduates of both sexes. The experiment was conducted in a soundproof room, partitioned by black curtains. The *S* was seated 5 ft. in front of a cutout in the curtains which permitted stimulus presentation at *S*'s eye level. The autokinetic stimulus was a red 37-w. bulb enclosed in a lightproof box. A pinpoint opening in the front of this box, .05 in. in diameter, permitted passage of the light. A constant speed motor provided for a continuous series of 6-sec. stimulus presentations and 9-sec. intertrial intervals. The stimulus could also be removed manually by means of a knife switch. A fan was used throughout the experimental sessions to mask *E*'s movements and to enhance autokinetic movement.

Procedure.—Each *S* was met outside the experimental room. He was told that the experiment would be carried out in a totally dark room. Upon entering the room he was asked to take his seat and keep his eyes closed. After 4 min. of dark-adaptation the following instructions were given to all *S*s.

"This is an experiment in visual discrimination. After I tell you what to do, I will turn on a red light which will appear in front of you. After a short while the light will start to move. Then it will go off. After the light is off I want you to tell me the distance the light has moved. Make your estimates in full inches, half-inches or quarter-inches. For example, say '5½ in.,' or '7¼ in.,' or '10½ in.,' and so on. Try to be as accurate as you can. The light will move no less than 2 in. and no more than 13 in.

"We found that people can judge more accurately when they hold still while the light is on. Also, please do not lean forward in your chair. There will be a short pause between lights and a short rest period after a while. We'll be here about 45 min. Do you have any questions?"

The first trial was begun after *S* located the stimulus and reported the onset and end of one presentation. Each *S* was given 30 trials in order to obtain a sample of his distribution of estimates. At the end of 30 trials *E* introduced a rest period of 1½ min. during which a *critical class of responses* for each *S* was chosen in the following manner: (*a*) Each *S*'s 30 estimates were arranged in order of numerical magnitude. (*b*) The lower limit of the critical class was set so that it exceeded the value of 70% or 21/30 of *S*'s initial responses. (*c*) The upper limit of the critical class was set so that it exceeded the numerical value of 90% or 27/30 of *S*'s initial responses. By this procedure the actual numbers which were included in the critical class varied for each *S* and were determined by the initial sample of 30 responses. On the basis of this sample, however, the probability of occurrence of a member of the critical class (henceforth called a critical response) was limited to a maximum value of .2 or 6/30 for any *S*.

Acquisition.—All *S*s had previously been assigned randomly to one of four groups, a 100%, 67%, 50%, or 0% reinforcement group. At the end of the rest period during which the critical class had been established for each *S*, acquisition was begun. Reinforcement consisted of three randomly arranged verbal responses by *E*: (*a*) "That's accurate," (*b*) "That's a correct estimate," and (*c*) "That's right on the button." Reinforcement was given immediately following each critical response in

TABLE 1

MEANS, *SD*'S, ADJUSTED MEANS, AND *F* RATIOS OF ACQUISITION MEASURES

Group	Trials			Reinforcement			Critical Responses	
	Mean	*SD*	Adj. Mean	Mean	*SD*	Adj. Mean	Mean	*SD*
100%	17.84	9.95	13.79	8.16	2.34	8.50	8.16	2.34
67%	28.16	14.52	26.58	7.68	2.01	7.43	10.48	3.20
50%	32.19	15.62	37.80	6.26	1.88	5.52	11.74	3.46
F ratio	8.89*			6.65*			10.78*	

* Significant at .01 level.

the 100% group. In the 67% group two of every three critical responses were reinforced, and the sequence of reinforcement (R's) was: RNRRNRRNR. In the 50% group every other critical response was reinforced. For these three groups acquisition ended when a criterion of six critical responses in ten consecutive trials was met. The probability of occurrence of the critical responses thus was raised from .2 to .6 in the 100%, 67%, and 50% groups. All *S*s in the 0% group were given 30 trials without reinforcement.

Extinction.—Seventy-five nonreinforced trials were given to all four groups, continuous with the acquisition procedure. At the end of the session each *S* was questioned concerning the experiment.

RESULTS

The 0% Reinforcement Group

The 0% group served as a control group to evaluate the effect of time alone in the autokinetic situation. A *t* test on the mean number of critical responses during Trials 31–60 (the acquisition trials in the other groups) and Trials 105–135 (the last 30 extinction trials in the reinforced groups) indicated that the frequency of critical responses did not change significantly as a function of time spent in the experiment ($t = .63$). This finding indicated the absence of any adaptation effect which might obscure the effects of the experimental variable. Data from this group were not included in the statistical treatments described below.

Acquisition

Number of trials to criterion.—The mean number of trials needed to reach the criterion of six critical responses in a block of ten trials is shown in Table 1. An analysis of variance showed that the differences between groups were significant at the .01 level of confidence. The number of trials increased with a decrease in the percentage of reinforcement. While this analysis clearly establishes the difference between the groups with respect to the number of trials, it does not attribute these differences to the varying percentages of reinforcement.

The presence of an uncontrolled number of reinforcements may underlie the differences in number of trials. An analysis of covariance was performed to equate for the effects of the number of reinforcements and a significant *F* (Table 2) was obtained. Adjustment of means for the groups showed that the direction of the difference was the same as for the nonadjusted data. The adjusted mean for the 100% group was 13.79; for the 67% group it was 26.58; and for the 50% group it was 37.80. Thus the difference in number of trials is due to other factors than the difference in number of reinforcements alone.

Number of reinforcements.—Table 1 indicates that a greater number of reinforcements had to be administered with the higher percentages of reinforcements in order to bring *Ss* to the same level of responding. In order to equate for the effects of the concomitant trial variable, an analysis of covariance was run and the group means were adjusted (Table 2). The significant difference ($F = 48.6$) in the same direction indicated that the different number of trials alone did not account for the difference in reinforcements among the groups. The adjusted means for the 100%, 67%, and 50% groups were: 8.50, 7.43, and 5.52, respectively.

Number of critical responses.—The *Ss* in the 50% group gave significantly more critical responses than *Ss* in the 67% and 100% group (Table 1). Since the 50% group had considerably more trials than the other groups, it might be supposed that the larger number of critical responses for the 50% group was due to the fact that *Ss* in that group had a greater opportunity to respond.

An analysis of covariance, adjusting for the number of trials, gave an *F* of 2.11 (Table 2) which is not significant at the .05 level of confidence. This analysis supported the position that the greater number of critical responses in the 50% group may be attributed to the greater number of trials required by *Ss* in that group. The effects of the concomitant reinforcement variable were tested by an analysis of covariance, adjusting for number of reinforcements (Table 2). A significant *F* ratio of 116.7 and adjustment of means indicated that the difference in critical responses cannot be attributed to the effect of the reinforcement variable alone.

In order to adjust the number of critical responses for both trials and reinforcements, a multiple covariance was run. The obtained *F* value of 13.8 (Table 2) pointed to an effect of the percentage-of-reinforcement variable on the number of critical re-

TABLE 2

ANALYSES OF COVARIANCE OF THE ACQUISITION DATA

Adjusted Variable	Source	*SS*	*df*	*F*
Reinforcements adjusted for trials	Between groups	172	2	48.6*
	Within groups	158	89	
	Total	330	91	
Critical responses adjusted for reinforcements	Between groups	490	2	116.7*
	Within groups	182	89	
	Total	672	91	
Critical responses adjusted for trials	Between groups	17	2	2.11
	Within groups	353	89	
	Total	370	91	
Trials adjusted for reinforcements	Between groups	7896	2	52.5*
	Within groups	6694	89	
	Total	14590	91	
Critical responses adjusted for trials and reinforcements	Between groups	240.2	2	13.8*
	Within groups	762.6	88	
	Total	1002.8	90	

* Significant at .01 level.

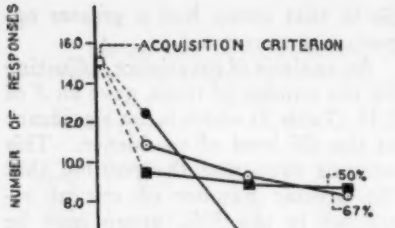


FIG. 1. Mean number of critical responses during extinction

sponses beyond the concomitant changes in trials and number of reinforcements alone.

Extinction

When over-all differences in the number of critical responses were compared (Table 3), an F ratio of 3.57, significant at the .05 level, was obtained. The 67% group gave the most, the 100% group the fewest critical responses. The curves in Fig. 1 show that the rate of extinction differs markedly for the groups, however. Furthermore, a reversal in the direction of the difference between partial and continuous groups occurred as extinction progressed. Therefore, a more meaningful analysis of the extinction data was attempted by measuring the rate of extinction rather than the total number of responses. For this purpose the ratio of Trials 1–25/51–75 for each S was computed. The mean value of this ratio reflected the slope of the extinction curve of each group. The mean ratios for the 100%, 67%, and 50% groups were 4.39, 1.64, and 1.30, respectively. Their SD's in the same order, were 3.56, 1.05, and .65. The direction of the difference between means and a significant F ratio (F = 18.13) suggested that partial reinforcement resulted in greater resistance to extinction than continuous reinforcement, when rate of deceleration was used as an extinction measure.

The group differences during extinction can also be analyzed in blocks of successive 25 trials. These data (Table 3) show that the rank order of groups changed as extinction progressed. During each block of trials F ratios on critical responses in that block indicate differences between groups. The direction of these differences changed as extinction progressed (Table 3). During the first trials,

TABLE 3

MEANS AND F RATIOS OF CRITICAL RESPONSES DURING EXTINCTION

Group	Trials 1–75		Trials 1–25		Trials 26–50		Trials 51–75	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
100%	22.52	8.38	12.45	4.59	5.90	3.30	4.13	2.66
67%	28.29	9.44	10.84	3.74	9.10	3.18	8.42	4.22
50%	27.03	8.56	9.48	3.11	9.03	4.23	8.52	2.91
F ratio	3.57*		4.47*		7.75**		16.90**	

* Significant at .05 level.
** Significant at .01 level.

for example, the 50% group gave the fewest critical responses, whereas during the last 25 trials it gave the most. The results of the analyses of the extinction data can be summarized as follows. Although the continuously reinforced group showed a high frequency of critical responses at the beginning of extinction, the frequency dropped rapidly as extinction proceeded. The partial groups[2] began with a lower frequency of critical responses than the continuous group, but showed slower deceleration of their response frequencies. In fact, as extinction progressed, the partial groups surpassed the continuous group in the number of responses given.

Effect of acquisition variables on extinction.—The experimental design provided for simultaneous operation of several variables during acquisition. Three of these concomitant variables (collectively called acquisition variables) were measured during the experiment. An attempt was made to isolate the effect of each of these variables on the total number of critical responses by analyses of covariance and multiple covariance. As has already been pointed out, extinction in this study consisted of a time sample which was limited because of practical considerations. Therefore, the following analyses reflect specific comparisons, restricted to the measure of extinction used in this study. A longer extinction period might have yielded different relationships.

The statistical control of each of the acquisition variables separately indicated that the differences which the groups showed on (*a*) number of reinforcements, (*b*) number of trials, and (*c*) number of critical responses during acquisition in each instance did not determine the group differences obtained in extinction. As indicated in Table 4, the *F* ratios increased in value after adjustment of each of the three acquisition variables.

TABLE 4

ANALYSES OF COVARIANCE OF THE EXTINCTION DATA

Adjusted Acquisition Variable	Source	*SS*	*df*	*F*
Trials	Between groups	870	2	5.66**
	Within groups	6832	89	
	Total	7702	91	
Reinforcements	Between groups	794	2	5.22**
	Within groups	6775	89	
	Total	7569	91	
Critical responses	Between groups	586	2	3.63*
	Within groups	7187	89	
	Total	7773	91	

* Significant at .05 level.
** Significant at .01 level.

A simultaneous adjustment was made for all three acquisition variables by means of a multiple analysis of covariance.[3] The obtained *F* of 70.11 with 2 and 84 *df* was significant at the .01 level of confidence. This means that additional variables in the acquisition procedure contributed to the difference between the groups during extinction after they were equated for the three measured acquisition variables. The *F* ratio, then, included the differences due to the percentage of reinforcement, possible pattern effects of the reinforcement schedule, and any residual variance due to deviation from linearity of the relationship among the three acquisition variables and the measures of extinction.

The answers to the postexperimental questions are of interest. Only 18 of the 104 *S*s stated that *E* had said anything at all during the experiment. Of these, 10 reported that *E* had made some statement concerning the ac-

[3] The method of solution by determinants was used. For derivation of the formulas see Wilks (12).

curacy of their estimates. Only 3 *Ss* could indicate the specific number range for which they had been reinforced. These findings suggest that mediating verbal responses or the verbal discrimination of the conditions for reinforcement by *Ss* are not necessary conditions for modification of verbal behavior in this situation.

Discussion

Acquisition.—Since resistance to extinction has been widely accepted as the major measure of the effects of varying reinforcement schedules, few studies have dealt primarily with acquisition. Some investigators, however, have reported that frequency of reinforcement had little effect on the number of trials in acquisition of a response, whereas others have found more rapid acquisition with continuous than with several schedules of partial reinforcement (9). A possible source of these discrepancies is brought out by the present experimental design. It has been found in the present study that the number of *trials* needed to reach a criterion is greater in partial groups while the number of *reinforcements* required is greater in continuous groups. In a design which holds constant the number of reinforcements, the number of critical responses given by the partial groups will exceed the number of critical responses in the continuous group. This is shown by the covariance data in Table 2. In experiments which employ a cumulative frequency curve of critical responses, the partial groups, therefore, will appear to excel the continuous groups when reinforcements are held constant. In experiments which employ a constant number of trials for all groups, a comparison of critical responses given by the partial and continuous groups may show no statistically significant differences, despite the advantages of continuous groups in the number of reinforcements.

This is indicated by the covariance analysis of critical responses, when the trial variable is adjusted (Table 2). However, the great advantage in the number of reinforcements in the continuous groups may result in the better performance at the end of a fixed number of trials. This suggestion is substantiated by the fact that the mean number of reinforcements, after adjustment for trials, was found to be 8.50 in the 100% group and 5.52 in the 50% group. With the number of trials experimentally fixed considerably below the number of trials required by the 50% group in the present study, the comparable figures would be in a ratio of better than 2:1 in favor of the 100% group. It is quite possible that the additional reinforcements administered under a design with fixed trials lead to the "slight superiority in acquisition" which Jenkins and Stanley find for the 100% groups (9). The relative efficiencies of partial and continuous reinforcement schedules in acquisition depend, therefore, on the specific measure in which the experimenter is interested and the purpose for which he wishes to employ a partial reinforcement schedule.

One recent study on partial reinforcement in a Thorndikian multiple-choice situation (1) reported data on several acquisition measures in a design comparable to the present one. Goss and Rabaioli report results which essentially agree with the present findings of more trials and fewer reinforcements in the 50% group and of an inverse relationship between mean number of critical responses and percentage of reinforcement during acquisition.

The complexity of the variables which operate in acquisition of a response under partial reinforcement is suggested by the significant F ratio after both trials and reinforcement had been equated in a multiple covariance. Among the additional variables requiring experimental investigation are the various components which comprise a "pattern of reinforcement." Exploration of some of these components has recently been reported by several authors (2, 5).

Extinction.—The present design attempted to equate the response strength of the experimental groups by means of a

criterion of performance. Nevertheless, the groups already showed differences during the first block of extinction trials (Fig. 1). A more detailed analysis of the data revealed the tendency of *S*s in the 100% group to exceed the criterion during the first ten nonreinforced trials. This tendency was infrequent in partially reinforced *S*s. The acquisition data suggest that the relative massing of reinforcements in the last ten trials of the continuous group, in comparison with the partial groups, may have resulted in the increase of critical responses immediately following cessation of reinforcement.

The use of the total number of critical responses in extinction as a measure of resistance to extinction has been shown to involve several difficulties. A more specific description of extinction progress is given by the rate of decline of critical responses and analysis of successive portions of the extinction curves. The results showed the fastest decline in the 100% group and the slowest in the 50% group. Similar sharp drops in 100% groups, compared with partial groups, have been found in eyelid conditioning (6), GSR conditioning (8), and verbal expectations (7). The significant *F* ratio in the multiple covariance, adjusting for the three acquisition variables, points to the need for additional research on the specific variables operating in acquisition. As the statistical analysis indicated, the three measured variables, taken singly or together, did not account for the differences between groups in extinction. It has already been noted that the "pattern of reinforcement" required further investigation. While similar differences between continuous and partial groups have been found in a variety of situations, consistent differences between partial groups, as a function of the percentage and pattern of reinforcement, have not yet been established. Another problem raised by the present study is the need for specificity with respect to response measures. Until precise relationships are established between variables which are said to define such constructs as "resistance to extinction" and "response strength," specification of the measures and conditions of the experiment must be made.

The present experimental design attempted to investigate a situation which involved some of the essential characteristics and relationships found in much of everyday human behavior. It employed a stimulus which tended to elicit, by instructions, a large variety of verbal responses. Also, the frequency of an arbitrarily defined response class was increased by selective (contingent) verbal reinforcement. Since the relationships among learning variables in the present study were found to be similar to those found in animal studies and classical conditioning situations (9), the results further support the universality of certain learning phenomena for different species and the continuity of nonverbal and verbal human behavior.

Summary

This study tested the effect of partial reinforcement on acquisition and extinction of verbal responses. The *S*s (124 undergraduates) were instructed to estimate movement of an autokinetic stimulus. A range of number responses at the upper end of each *S*'s distribution of 30 initial estimates was selected for reinforcement. The *S*s were divided into 100%, 67%, 50%, and 0% reinforcement groups. During acquisition each *S*'s critical responses were followed by verbal reinforcement, according to the schedule of his group. After a criterion of six critical responses in ten consecutive responses was reached, extinction was carried on for 75 nonreinforced trials for all experimental groups.

It was found that: (*a*) Acquisition with a partial reinforcement procedure requires more trials but fewer reinforcements than acquisition with a continuous reinforcement procedure, and (*b*) extinction with a continuous reinforcement procedure shows greater initial resistance but faster decline than extinction with a partial reinforcement procedure. During extinction, the total number of critical responses was found to be greatest in the 67% group, although the 50% showed the slowest deceleration in responses.

The results of this study were compared with findings in classical human conditioning and in animal studies. The difficulties of experimental control in partial reinforcement studies were

pointed out and statistical methods of control which were used in this study were discussed.

References

1. Goss, A. E., & Rabaioli, E. J. Response-strength in a modified Thorndikian multiple-choice situation as a function of varying proportions of reinforcement. *J. exp. Psychol.*, 1952, **43**, 106–113.
2. Grant, D. A., Riopelle, A. I., & Hake, H. W. Resistance to extinction and the pattern of reinforcement. I. Alternation of reinforcement and the conditioned eyelid response. *J. exp. Psychol.*, 1950, **40**, 53–60.
3. Greenspoon, J. The effect of verbal and non-verbal stimuli on the frequency of members of two verbal response classes. Unpublished doctor's dissertation, Indiana Univ., 1951.
4. Haggard, E. A., & Rabin, R. On the problem of 'reinforcement' in conditioning the autokinetic phenomenon. *J. exp. Psychol.*, 1948, **38**, 511–525.
5. Hake, H. W., Grant, D. A., & Hornseth, J. P. Resistance to extinction and the pattern of reinforcement. III. The effect of trial patterning on verbal "conditioning." *J. exp. Psychol.*, 1951, **41**, 221–225.
6. Humphreys, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.
7. Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
8. Humphreys, L. G. Extinction of conditioned psycho-galvanic responses following two conditions of reinforcement. *J. exp. Psychol.*, 1940, **27**, 71–75.
9. Jenkins, W. O., & Stanley, J. C. Partial reinforcement. *Psychol. Bull.*, 1950, **47**, 193–234.
10. McGeoch, J. A. *The psychology of human learning*. New York: Longmans, Green, 1942.
11. Skinner, B. F. *The behavior of organisms*. New York: Appleton-Century, 1938.
12. Wilks, S. S. *Mathematical statistics*. Princeton, N. J.: Princeton Univ. Press, 1943.

(Received March 31, 1954)

Journal of Experimental Psychology
Vol. 48, No. 6, 1954

STIMULUS GENERALIZATION AS A FUNCTION OF VERBAL REINFORCEMENT COMBINATIONS[1]

ARNOLD H. BUSS, MORTON WEINER, AND EDITH BUSS

Carter Memorial Hospital, Indianapolis

The variables of verbal reinforcement and nonreinforcement in human learning have been receiving more attention in recent years. In particular the functional relationships between *E*'s verbalizations and *S*'s learning have been the focus of several studies. Greenspoon (**4**) demonstrated the effectiveness of *um-hum* as a reinforcer of college student's verbal responses. Taffel (**5**) showed that *good* facilitated learning of certain verbal responses by psychiatric patients. In both studies reinforcement was given after the correct verbal response was elicited, but *E* said nothing after an incorrect response. Thus the reinforcement combination was *Um-hum*-Nothing in Greenspoon's study and *Good*-Nothing in Taffel's study.

In an experiment by Garmezy (**3**) two reinforcement combinations were used in a situation involving generalization from one tone to others on the pitch continuum. First, psychiatric patients and normals were given a *Right*-Nothing combination; then, after stressful instructions were introduced, the patients and normals were given a *Right-Wrong* combination. It was found that with normals the *Right-Wrong* combination resulted in a steeper generalization gradient than *Right*-Nothing. With the psychiatric patients the generalization gradients had similar slopes for the two reinforcement combinations. Garmezy concluded that the punishing effects of *Wrong* prevented the patients from acquiring differential response tendencies, with the result that their gradients remained relatively flat. However, it should be noted that *S*s receiving the *Right-Wrong* schedule also were given stressful instructions. Since the variables were confounded, it is not clear whether the difference in performance between normals and patients may be attributed to reinforcement combinations or to stress. In addition, Garmezy's experimental situation appears to be similar in some ways to the discrimination learning paradigm. During the test series he reinforced and nonreinforced *S*'s responses so that further learning was occurring while generalization was being tested.

It was the purpose of the present study to investigate the effect of similar verbal reinforcement combinations (in the absence of stress) upon primary stimulus generalization gradients.

METHOD

Materials.—The stimuli, described elsewhere (**1**), were wooden discs of various colors, areas, and heights. There were five colors, three areas, and four heights. The heights were 2 in., 1 in., .5 in., and .25 in. Every stimulus was different from every other stimulus with respect to color, height, or area. The various colors and areas were divided equally among the four heights.

Procedure.—The *E* read the following instructions: "I am going to show you a series of wooden blocks, one at a time. Some of the blocks are *Vec*, (V-E-C), and some are *not-Vec*. You don't know what a *Vec* is now, but it will become clear to you as we go. To begin with you may have to guess." Each stimulus was presented singly for approximately 5 sec. After each response *E* gave the appropriate verbal reinforce-

[1] This study was carried out with the permission and the encouragement of Juul Nielsen, M.D. The writers are also indebted to Dr. W. K. Estes for his suggestions in planning and interpreting the results of this study.

ment or nonreinforcement, and then the next stimulus was presented.

At the end of the first series *E* read the following: "Now we'll keep going just the way we are, except that I will no longer say anything after you respond." This instruction was given to prevent extinction during Series 2, which served as a test series.

Design.—There were two series of stimuli, a learning series and a test series. Series 1 consisted of eight 2-in. blocks, and *Vec* equalled 2-in. blocks. All *Vec* responses were correct, and all *not-Vec* responses were incorrect. The verbal reinforcement combinations given by *E* were as follows: Group I—*Right* after a *Vec* response, *Wrong* after a *not-Vec* response; Group II—nothing after a *Vec* response, *Wrong* after a *not-Vec* response; and Group III—*Right* after a *Vec* response, nothing after a *not-Vec* response.

Series 2 was the same for all *Ss*. It consisted of 11 each of 2 in., 1 in., .5 in., and .25 in. high blocks.

Subjects.—The *Ss* were 102 patients in a neuropsychiatric hospital who were not on treatment and who were not extremely disturbed. Patients who failed to make both responses (*Vec* and *not-Vec*) in the first series were excluded from the study. It was essential that both reinforcement and nonreinforcement occur in Series 1 or the particular reinforcement combination would not be effective. For example, 16 *Ss* originally assigned to Group III (*Right*-Nothing) responded *Vec* to all the stimuli in Series 1. They received only the *Right* component of the *Right*-Nothing combination, and the experimental condition was not met. Therefore, these *Ss* were excluded from the study. The following 26 *Ss* were also excluded from the study: 9 *Ss* originally assigned to Group I (*Right-Wrong*) who responded *Vec* to every stimulus and never heard *E* say *Wrong*; 9 *Ss* originally assigned to Group II (Nothing-*Wrong*) who responded *Vec* to every stimulus and to whom *E* said nothing; and 8 *Ss* originally assigned to Group III who responded *not Vec* to every stimulus and to whom *E* said nothing. Thus, 42 *Ss* were discarded for failing to make both responses in Series 1, leaving three groups of 20 *Ss*.

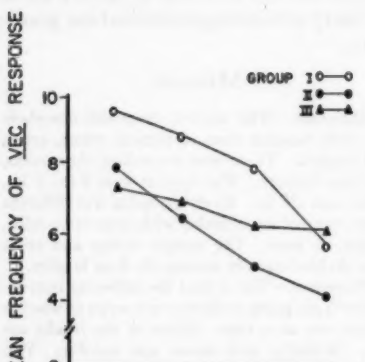


FIG. 1. Primary stimulus generalization gradients in the test series

TABLE 1

ANALYSIS OF VARIANCE OF *Vec* RESPONSE FREQUENCIES IN GROUPS I, II, AND III IN SERIES 2

Source	*df*	Mean Square	*F*
Between groups	2	105.83	3.18*
Between *Ss* in same group	56	33.23	
Between stimuli	3	80.94	17.19**
Groups × Stimuli	6	10.94	2.42*
Pooled *Ss* × Stimuli	171	5.42	
Total	239		

* $p = .05$.
** $p = .001$.

RESULTS

The measure of response strength used here is the frequency of the *Vec* response[2] in the test series. The major results are presented in Fig. 1. The slopes of the generalization gradients differ among the groups. The generalization gradient of Group III is flatter than the gradients of Groups I and II. There are also differences in the over-all height of the gradients; in the descending order of *Vec* response frequency are Groups I, III, and II. The significance of these trends was tested by an analysis of variance for repeated measures on the same *Ss* (2), which is presented in Table 1. This analysis indicates that (*a*) the groups differ significantly in their over-all *Vec* response frequency, (*b*) there are significant differences in response frequencies to the various stimuli on the height continuum, and (*c*) the slopes

[2] The frequency of the *Vec* response is inversely proportional to the frequency of the *not-Vec* response. Therefore, only the frequency of the *Vec* response will be considered.

of the generalization gradients are significantly different among groups. Since there were significant differences among groups in their over-all *Vec* response frequency, t tests were performed to test the significance of the individual comparisons. The results were: Group I vs. Group II, $t = 2.39$, $p < .05$; Group I vs. Group III, $t = 1.74$, $p < .10$; and Group II vs. Group III, $t = .84$.

Discussion

The results indicate that, in contrast to Garmezy's findings, a *Right-Wrong* reinforcement combination with psychiatric patients yields a steeper generalization gradient than does *Right*-Nothing. In evaluating this discrepancy in results, two differences in experimental procedures may be noted. First, in Garmezy's test series correct responses were reinforced and incorrect responses were nonreinforced. This differential reinforcement should have steepened the generalization gradients. However, the *S*s' generalization gradients remained flat, which suggests that the first difference in procedure is relatively unimportant.

The second difference was the confounding of stress with reinforcement combinations in Garmezy's study, whereas in the present study there was no stress and therefore no confounding. In Garmezy's experiment stress was present with the *Right-Wrong* condition but not the *Right*-Nothing. His normal *S*s had steeper generalization gradients with *Right-Wrong* than with *Right*-Nothing. His patient *S*s failed to manifest this difference in the slope of generalization gradients. Is the difference in performance between patients and normals due to stress or to reinforcement combinations? The present study would seem to provide an answer. The *S*s here were not stressed, and they performed like Garmezy's normals, i.e., *Right-Wrong* resulted in a steeper generalization gradient than *Right*-Nothing. If it is assumed that normals and patients react differently to stress, the difference in performance between Garmezy's patients and normals may be attributed solely to stress.

In attempting to account for the results of the present study several assumptions will be made concerning the reinforcing properties of *E*'s verbalizations. When a response is followed by *Right*, the probability of the response occurring again may be expected to increase, i.e., the response has been reinforced. When a response has been followed by *Wrong*, the probability of the response occurring again may be expected to decrease; i.e., the response has been nonreinforced. When a response has been followed by nothing, the probability of the response occurring again may be expected to remain unchanged (in human learning situations), i.e., the response has been neither reinforced nor nonreinforced. These assumptions may help account for the results of this experiment.

In Series 2 the *Right-Wrong* group had a higher over-all *Vec* response frequency than either of the other two groups. The difference between *Right-Wrong* and Nothing-*Wrong* was significant, and the difference between *Right-Wrong* and *Right*-Nothing approached significance. The explanation for these differences requires an examination of the learning situation in Series 1.

In the learning series all *Vec* responses were correct. In the *Right-Wrong* and *Right*-Nothing groups *E* said *Right* after a *Vec* response, which should have increased the *Vec* response strength. In the Nothing-*Wrong* group *E* said nothing after a *Vec* response, which should have had no effect on the *Vec* response strength. Thus, examination of *Vec* responses alone suggests that the *Right-Wrong* and *Right*-Nothing groups should have a higher *Vec* response frequency in the test series than the Nothing-*Wrong* group.

Examination of both *Vec* and *not-Vec* responses reveals that (*a*) in the *Right-Wrong* group the *Vec* response strength would be increased by both reinforcement of *Vec* and nonreinforcement of *not Vec;* (*b*) in the *Right*-Nothing group the *Vec* response strength would be increased by only reinforcement of the *Vec* response; and (*c*) in the Nothing-*Wrong* group the

Vec response strength would be increased by only nonreinforcement of the *not-Vec* response. On the basis of these considerations the over-all *Vec* response frequency of the *Right-Wrong* group would be expected to be higher than that of the other two groups, an expectation that conforms with the results of the experiment.

The generalization gradients in the test series may be analyzed in a similar fashion. In the *Right-Wrong* and *Right*-Nothing groups reinforcement of the *Vec* response to 2-in. stimuli in the learning series should increase the response frequency to 2-in. stimuli. The decreasing *Vec* response frequency to shorter stimuli in the test series would form a gradient of primary stimulus generalization. In the Nothing-*Wrong* group the absence of either reinforcement or nonreinforcement of the *Vec* response in Series 1 should result in a flat generalization gradient in Series 2. On the other hand nonreinforcement of the *not-Vec* response to 2-in. stimuli in the *Right-Wrong* and Nothing-*Wrong* groups should decrease its frequency to 2-in. stimuli. In this two-response situation a decrement in *not-Vec* response strength must result in an increment in *Vec* response strength. This *Vec* response tendency to 2-in. stimuli would then generalize to shorter stimuli, forming a stimulus generalization gradient. In the *Right*-Nothing group, however, the frequency of the *not-Vec* response to 2-in. stimuli would be unaffected, and a flat generalization gradient would result. Examination of both *Vec* and *not-Vec* responses suggests that the *Right-Wrong* group should have a steeper generalization gradient than those of the other two groups. The expectation is not borne out by the results, for both the *Right-Wrong* and the Nothing-*Wrong* groups had steeper gradients than the *Right*-Nothing group.

This discrepancy can be explained if it is assumed that saying nothing after a response is a very mild reinforcer. If Nothing had slight reinforcing properties,[3] the Nothing-*Wrong* group would approach the *Right-Wrong* group as a reinforcement combination. On this basis the steepness of the Nothing-*Wrong* generalization gradient should approach that of the *Right-Wrong* gradient. In the *Right*-Nothing group the difference between *Right* and Nothing would be less than expected, and the generalization gradient would tend to remain flat, as may be seen in Fig. 1.

Summary

An experiment was performed in order to investigate the effect of verbal reinforcement combinations on stimulus generalization, employing neuropsychiatric patients as *S*s. There were three reinforcement combinations: *Right-Wrong*, Nothing-*Wrong*, and *Right*-Nothing. The stimuli were wooden discs and the generalization continuum was height. The *Right-Wrong* and Nothing-*Wrong* combinations resulted in significantly steeper gradients of primary stimulus generalization than the *Right*-Nothing combination. There were also significant differences among groups in the height of the generalization gradients.

These findings were accounted for in terms of the reinforcing properties of *E*'s verbalizations, and the results were related to previous studies in this area.

References

1. Buss, A. H. Rigidity as a function of absolute and relational shifts in the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45**, 153–156.
2. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
3. Garmezy, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *J. Pers.*, 1951, **20**, 253–276.
4. Greenspoon, J. The effect of verbal and non-verbal stimuli on the frequency of members of two verbal response classes. Unpublished doctor's dissertation, Indiana Univ., 1951.
5. Taffel, C. Conditioning of verbal behavior in an institutionalized population and its relation to anxiety level. Unpublished doctor's dissertation, Indiana Univ., 1952.

(Received April 1, 1954)

[3] The *E*'s saying nothing after a response is not expected to have any reinforcing properties except in the context of differential reinforcement, e.g., *Right*-Nothing or Nothing-*Wrong*. If *E* were never to say anything after *S*'s responses, no learning would occur.

Journal of Experimental Psychology
Vol. 48, No. 6, 1954

DIMENSIONAL ANALYSIS OF PSYCHOMOTOR ABILITIES[1]

EDWIN A. FLEISHMAN

Skill Components Research Laboratory, Air Force Personnel and Training Research Center

The present investigation is concerned with the isolation and definition of important dimensions of psychomotor skill. It represents one in a series of studies (e.g., **1, 6, 7, 8, 9, 11**), the aim of which is to provide a functional classification of abilities adequate to account for individual differences in psychomotor performance. In other terms, what we are seeking is a limited number of ability categories which might be useful and meaningful in describing performance in a wide variety of psychomotor tasks.

These studies were originally undertaken within the framework of aptitude measurement research in the psychomotor area. It was felt that a fundamental need in this area of measurement was an extensive dimensional analysis of psychomotor abilities which would serve as a basis for aptitude test development in this area.[2] It would appear, however, that research along these lines also has relevance for research in experimental psychology in a more general sense. Knowledge about the organization of abilities in the psychomotor area is, of course, important in its own right. However, such knowledge may also serve to integrate laboratory research on psychomotor skills where the primary interest is in the relationships among task variables and individual differences variables on the one hand, and procedural or other environmental variables on the other.

The primary approach in the present series of studies is to observe covariation in the performance of individuals on a large variety of especially designed psychomotor tasks. For the most part, in any particular study, these tasks are constructed with a view to certain definite hypotheses about the organization of abilities contributing to performance on these tasks. It then remains to verify or modify these hypotheses through analysis of the correlation patterns obtained on these tasks. Specifically, the techniques of factor analysis are applied to the intercorrelations in order to delineate more precisely the possible sources of common variation underlying these performances. It may be stressed that there are obviously other ways of categorizing and describing such skills. However, it is felt that factor analysis provides the best available technique for classifying skills into the smallest number and yet most independent set of categories that might describe performance in a wide variety of tasks. It would also appear that a dimensional analysis derived empirically from the intercorrelations of performances provides a classification that is more functional from the point of view of the responding individual.

Research of this kind in the psychomotor area has been fragmentary and for the most part has investigated only a limited range of psychomotor tasks.

[1] This research was carried out under the Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas, in support of Project 7703. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government. This study has been reported in somewhat different form in an AFPTRC Research Bulletin 54-15 (**2**).

[2] The writer has discussed the implications of this approach in detail in a previous paper (**4**).

However, a number of psychomotor factors have been identified independently by several different investigators. The writer has presented a review of such previous studies elsewhere (**4**). The purpose of the present report is to describe the development and factor analysis of a battery of 38 specifically designed apparatus and printed psychomotor tests. This study was an attempt to verify empirically in a single study practically all of the psychomotor factors previously identified in the separate studies, together with any new factors that might emerge. More specifically, it was felt that a factor analysis of the intercorrelations among these tests would indicate (*a*) the tenability of the factors hypothesized or the presence of new factors which might better account for the interrelationships, (*b*) the possibility of breaking down a general factor of psychomotor coordination (previously found in certain complex apparatus tests) into more basic ability categories, (*c*) the extent to which more complex psychomotor performance could be accounted for by the factors identified, and (*d*) the utility of specifically designed printed psychomotor tests in reproducing factor variance involved in their apparatus test counterparts.

Hypothesized Factors

The tests were constructed with a view to certain factors hypothesized from the results of the previous review of the literature (**4**). These factors and their tentative definitions are as follows:

1. *Reaction Time.*—The speed with which an individual is able to respond, by means of a prescribed movement, to a stimulus when it appears.

2. *Speed of Simple Ballistic Movement (Tapping).*—The speed with which an individual is able to oscillate either his finger or his arm, independent of any eye-hand coordination ability.

3. *Manual Dexterity.*—The ability to make skillful, controlled arm and hand manipulations at a rapid rate.

4. *Finger Dexterity.*—The ability to make skillful, controlled manipulations with the fingers at a rapid rate.

5. *Precision of Movement (Steadiness).*—The precision and steadiness with which one is able to make accurate arm-hand positioning movements of the type which minimize strength and speed.

6. *Aiming.*—The ability to carry out quickly *and* precisely a series of accurately directed movements requiring eye-hand coordination.

7. *Motor-Kinesthesis.*—The ability to make precise postural adjustments to kinesthetic cues when the body or body members are displaced from positions of equilibrium.

In addition, tests were included in the battery which had been identified in previous analyses of the Air Force's Aircrew Battery as sampling two additional tentative factors—*Psychomotor Precision* and *Psychomotor Speed.* It was hoped that the present analysis might clarify the status of these less well-defined factors.

At least three tests were constructed aimed at sampling each ability category. In addition to these tests, the operational Air Force psychomotor tests were added to the battery. It was hoped that this would throw some light on the nature of the *Psychomotor Coordination* factor previously found common to the psychomotor aircrew tests in repeated factor analyses of the Aircrew Battery. It was hypothesized that this general factor of Psychomotor Coordination could be broken down into more basic factors of psychomotor skill. Heretofore, this had not been possible since previous factor analysis batteries which have included the Aircrew Tests have not included other psychomotor tests aimed at reproducing variance in the Air Force psychomotor tests.

Descriptions of the Tests

Brief descriptions of the tests follow. While most of the experimental tests are new, a few have been adapted from existing tests. In addition, certain commercially available tests

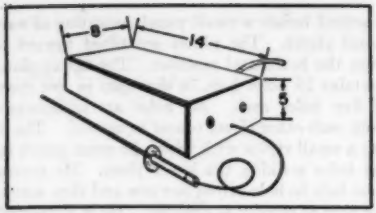


FIG. 1. PRECISION – STEADINESS

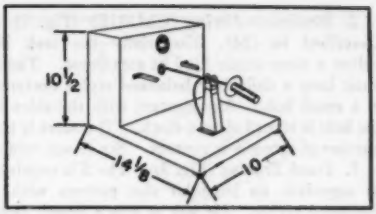


FIG. 2. STEADINESS – AIMING

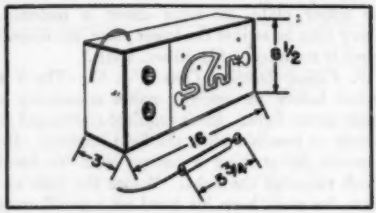


FIG. 3. TRACK TRACING

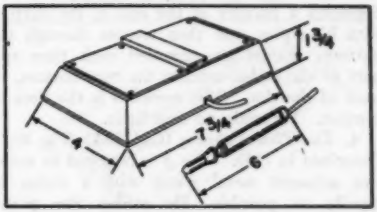


FIG. 4. TWO-PLATE TAPPING

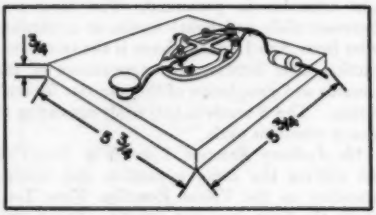


FIG. 5. KEY TAPPING

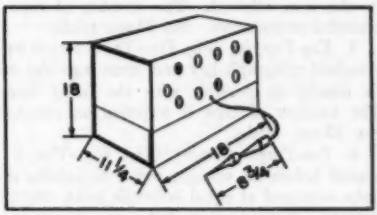


FIG. 6. TEN TARGET AIMING

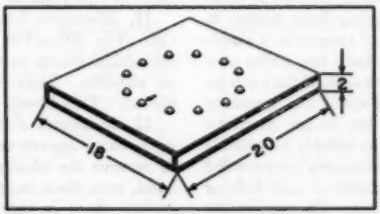


FIG. 7. ROTARY AIMING

were included. Some of the printed tests are largely adaptations of tests similar to those in the MacQuarrie Tests of Mechanical Ability (13). The apparatus tests are pictured in Fig. 1 to 21. All of the dimensions drawn in each of these figures are in terms of inches.

Apparatus Tests

1. *Precision-Steadiness* (Fig. 1).—The *S* is seated before a long rectangular boxlike apparatus containing two openings. Each opening is the entrance to a straight passageway which *S* must negotiate with a long stylus. He moves the stylus *forward* at slightly below shoulder height and at arm's length. He must move the stylus slowly and steadily away from his body, trying not to hit the sides of the cylindrical passage. As he reaches the end of the passage he strikes a contact point and withdraws the stylus, again trying to avoid hitting any part of the passageway. He then negotiates the second passageway. Two complete negotiations constitute a trial. Counters record the number of contacts and clocks record the amount of time in contact. Six trials, no time limit.

2. *Steadiness-Aiming* (CM103E) (Fig. 2).—Described in (14). Essentially the task involves a more static kind of steadiness. The *S* must keep a delicately balanced stylus centered in a small hole. Any contact with the sides of the hole is recorded on a clock. The score is the number of seconds in contact. Six 40-sec. trials.

3. *Track Tracing* (Fig. 3).—The *S* is required to negotiate an irregular slot pattern with a T-shaped stylus. He sits at arm's length from the apparatus box and moves slowly and steadily through the pattern from right to left, depresses a plunger at the end of the pattern with his stylus, and then returns through the pattern. Errors are recorded each time any part of the stylus touches the top, bottom, or back of the slot. Also recorded is the time in contact. Six trials, no time limit.

4. *Two-Plate Tapping* (CM202A) (Fig. 4).—Described in (14). The *S* is required to strike two adjacent metal plates with a stylus as rapidly as possible. He strikes the plates successively; that is, first one then the other, making as many taps as possible on the plates in the time allowed. The number of taps is recorded on counters. Six 30-sec. trials.

5. *Key Tapping* (Fig. 5).—The *S* stands by a standard telegraph key and must tap the key as rapidly as possible with the index finger. The number of taps is recorded on counters. Six 30-sec. trials.

6. *Ten-Target Aiming* (Fig. 6).—The *S* is seated before an upright panel containing ten holes arranged at equal intervals in an elliptoid pattern. Behind each hole can be seen a circular target. These targets vary in size from hole to hole. The *S* is required to strike at these targets with a stylus, moving from target to target around the pattern of targets in a clockwise direction. He makes only one strike at a time in each hole as he moves around the pattern. He is instructed that both speed and accuracy count and that he must try to hit as many targets as possible, moving as quickly as possible from target to target. Error counts are recorded each time *S* strikes the outside of each hole or inside around the target area. Correct counts are scored each time *S* hits within the target area in each hole. Six 30-sec. trials.

7. *Rotary Aiming* (Fig. 7).—The *S* stands before a horizontal panel containing 12 buttons which extend ¼ in. above the panel and are arranged in a circular pattern. The *S* is required to strike each button with the index finger, moving from button to button in a clockwise direction. He makes only one strike at each button each time as he continues around the circular pattern, as rapidly as possible, until told to stop. Score is the number of strikes in the time allowed. Six 30-sec. trials.

8. *Hand-Precision Aiming* (Fig. 8).—The *S* is seated before a small panel consisting of two metal plates. The plates are tilted toward *S* from the horizontal position. The upper plate contains 25 holes ¼ in. in diameter in five rows of five holes each. All holes are equidistant from each other (from center to center). The *S* has a small stylus with which he must punch in the holes striking the lower plate. He moves from hole to hole across one row and then across the next as rapidly as possible. He is instructed to aim accurately with each punch but to work as rapidly as possible. Every time he strikes the upper plate, an error count is recorded. Every time he strikes the lower plate, the correct count is recorded. Six 30-sec. trials.

9. *Visual Reaction Time* (Fig. 9).—The *S* is seated before an upright panel containing a single amber light. He is required to respond as rapidly as possible when the light appears. He responds by striking a button with his hand which turns off the light. Before the light appears, he must keep his hand on a small cross located 6 in. in front of the button. A click provides him with a ready signal before each light stimulus is presented. The foreperiod (between click and light) varies in a random order from .5 to 1.5 sec. Score is the cumulated reaction time between the appearance of the stimulus and completion of the response for each setting. The *S* receives two trials consisting of twenty reactions each.

10. *Auditory Reaction Time* (Fig. 9).—This test utilizes the same apparatus and scoring procedure as the Visual Reaction Time Test, except the light does not appear. The *S* now responds to a buzzer when it sounds. The *S* receives two trials of 20 reactions each.

11. *Minnesota Rate of Manipulation—Placing* (15) (Fig. 10).—The *S* is required to place 60 cylindrical blocks in the proper holes as rapidly as possible. Score is the number of blocks placed. Two 40-sec. trials.

12. *Minnesota Rate of Manipulation—Turning.*—Same apparatus as 11. The *S* is required to remove the blocks from the holes with one hand, turn them over with the other hand, and replace them in the same holes, moving from block to block as rapidly as possible. Score is the number of blocks turned. Two 35-sec. trials.

13. *Purdue Pegboard—Right Hand* (16) (Fig. 11).—The *S* is required to place a number of small pegs individually in a series of small holes as rapidly as possible with the right hand. Score is the number of pegs placed. One 30-sec. trial.

14. *Purdue Pegboard—Left Hand.*—Same as 13 with the left hand.

15. *Purdue Pegboard—Both Hands.*—The *S* is required to pick up two pins at a time, one with each hand from different trays and place

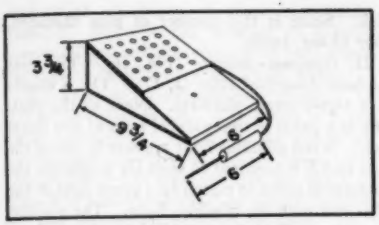


FIG. 8. HAND-PRECISION AIMING

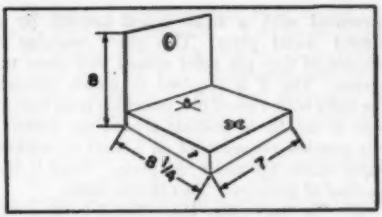


FIG. 9. VISUAL REACTION TIME
AUDITORY REACTION TIME

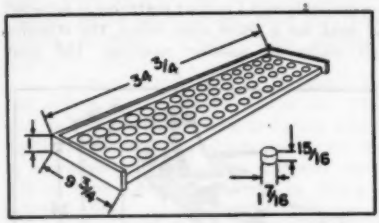


FIG. 10. MINNESOTA RATE OF MANIPULATION

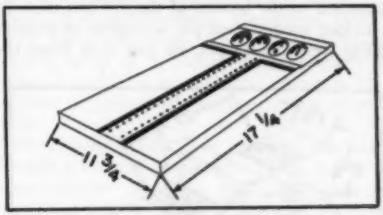


FIG. 11. PURDUE PEGBOARD

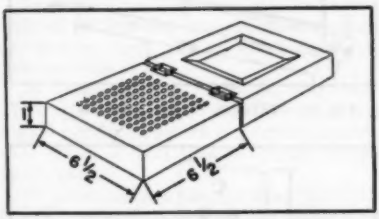


FIG. 12. O'CONNOR FINGER DEXTERITY

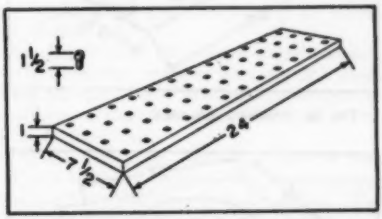


FIG. 13. SANTA ANA FINGER DEXTERITY

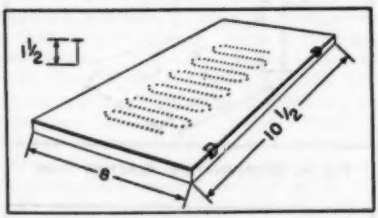


FIG. 14. PUNCH BOARD

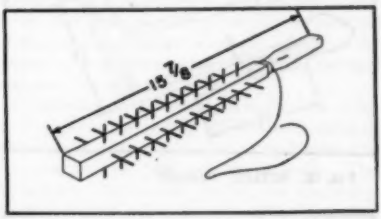


FIG. 15. PIN-STICK

them simultaneously in two different holes. Score is the number of pegs placed. One 30-sec. trial.

16. *Purdue Pegboard—Assembly.*—The *S* is required to make as many completed peg-washer-collar-washer assemblies as possible in the time allowed. Score is the number of assembly components completed. One 60-sec. trial.

17. *O'Connor Finger Dexterity* (12) (Fig. 12).—The *S* is required to pick up three small pins at a time from a tray of pins with the preferred hand and place them three at a time in a small hole. He must fill a series of small holes in this manner as fast as possible. Score is the number of pins placed. One 5-min. trial.

18. *Santa Ana Finger Dexterity* (CM116A) (Fig. 13).—Described in (13). The *S* is required to rotate a number of square pegs (with circular tops) 180° in their holes moving from peg to peg as rapidly as possible. Score is the number of pegs rotated. Two 35-sec. trials.

19. *Punch Board* (Fig. 14).—The *S* is

presented with a small board covered by a hinged metal plate. This plate contains a pattern of tiny pin holes spaced very close together. The *S* is required to punch through the holes with a small pin, punching from hole to hole as rapidly as possible around the pattern. His punches are recorded on a sheet of marked paper which fits under the plate. Score is the number of punches. Two 60-sec. trials.

20. *Pin Stick* (Fig. 15).—The *S* holds a rod containing four rows of pins on each of four sides. He is required to take the thread attached to the bottom of the rod and to make one loop around each pin as rapidly as possible going from pin to pin, up and then down the stick. Score is the number of pins threaded. Four 15-sec. trials.

21. *Dynamic Balance* (Fig. 16).—This test has been described fully (**3, 17**). The *S* stands on a teeter-totter platform. Next to the platform is a panel of two parallel rows of five lights each. A red stimulus light appears in one of the rows and *S* is required to shift his weight on the platform in order to match up a green light in the other row with the stimulus light. The position of the green light is controlled by the position of the platform. When the platform is tilted appropriately and a correct matching is achieved and held for a short time delay, the stimulus light shifts to another position. He must

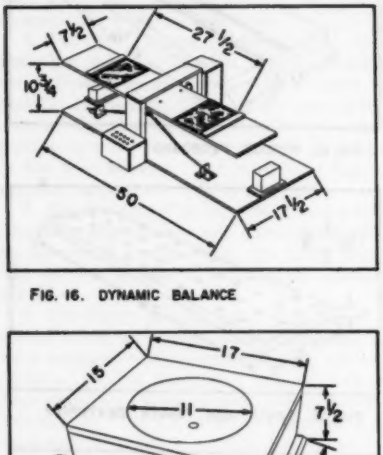


FIG. 16. DYNAMIC BALANCE

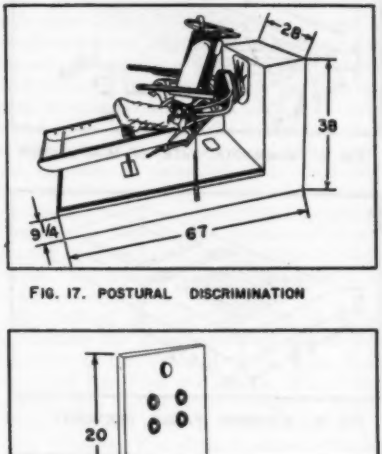


FIG. 17. POSTURAL DISCRIMINATION

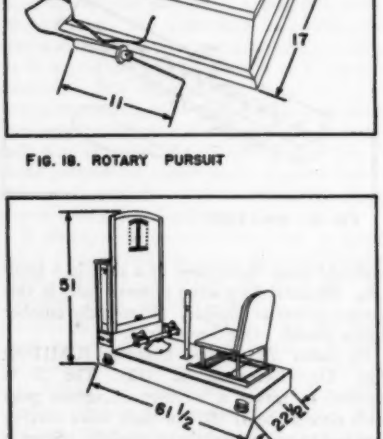


FIG. 18. ROTARY PURSUIT

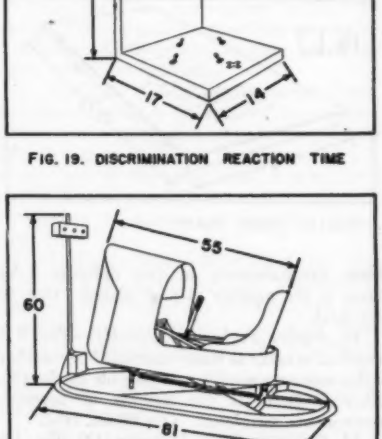


FIG. 19. DISCRIMINATION REACTION TIME

FIG. 20. COMPLEX COORDINATION

FIG. 21. RUDDER CONTROL

accomplish as many of these matchings as possible. Score is the number of matchings accomplished and the amount of time spent in the correct positions. Four 2-min. trials.

22. *Postural Discrimination—Vertical* (Fig. 17).—Described in (5). The *S* sits in a tilting chair arrangement which can be displaced to various positions by the administrator. The *S* is blindfolded and strapped into the chair. His task is to return the chair as closely as possible to the perceived upright position. He readjusts the chair by means of buttons on the arm rests. Score is the average deviation (number of degrees) of *S*'s readjustment from the "true" upright position. Twelve trials (displacements), no time limit.

23. *Postural Discrimination—Angular.*—This test utilizes the same apparatus as 22. However, this time *S* is displaced to a given angular position, held in that position briefly, and told to remember it. He is then displaced and is told to reproduce that position as closely as possible. Score is the average deviation (in degrees) from the correct position. Twelve trials (displacements), no time limit.

In addition to these tests, the final battery included the following four Air Force apparatus psychomotor tests, all of which have been described fully in (14).

24. *Rotary Pursuit* (CM803B) (Fig. 18).—The *S* attempts to keep a stylus in contact with a small metallic target set in a rapidly revolving disk. Score is total time on target. Fifteen 20-sec. trials.

25. *Discrimination Reaction Time* (CP611D2) (Fig. 19).—The *S* manipulates one of four toggle switches as rapidly as possible in response to a series of visual patterns differing from one another with respect to the spatial arrangement of their component parts. Score is cumulated reaction time. Four trials of 20 reactions each.

26. *Complex Coordination* (CM701E) (Fig. 20).—The *S* is required to make complex motor adjustment of stick and pedal controls in response to successively presented patterns of visual signals. Score is the number of completed matchings. One 8-min. test period.

27. *Rudder Control* (CM120C) (Fig. 21).—The *S* sits in a mock cockpit which his own weight throws off balance unless he applies proper correction by means of foot pedals. His task is to keep the cockpit lined up with one of three target lights. Score is total time on target. Three 30-sec. center target trials and three 112-sec. triple target trials.

Printed Tests

28. *Medium Tapping.*—The *S* is required to make three dots in each of a series of circles $\frac{1}{4}$ in. in diameter, working as rapidly as possible. Two 15-sec. trials.

29. *Large Tapping.*—Same as 28 except the circles are $\frac{1}{2}$ in. in diameter. Two 60-sec. trials.

30. *Aiming.*—The *S* is required to make one dot in a series of very small circles ($\frac{1}{8}$ in. in diameter), working as fast and accurately as possible. Score is number of dots correctly placed. Two 30-sec. trials.

31. *Pursuit Aiming—I.*—The *S* is required to follow a pattern of small circles ($\frac{1}{16}$ in. in diameter) placing one dot in each circle around the pattern. Two 15-sec. trials.

32. *Pursuit Aiming—II.*—Same as 31 except the pattern is more difficult and the circles are smaller ($\frac{1}{8}$ in. in diameter). Two 60-sec. trials.

33. *Square Marking.*—The *S* is required to place a series of × marks precisely inside a series of small ($\frac{1}{4}$-in.) squares. Score is the number of completed squares. Two 60-sec. trials.

34. *Tracing.*—The *S* is required to trace through a series of small openings ($\frac{1}{16}$ in.) in a maze pattern. He must work as quickly as possible trying not to allow his pencil mark to touch any of the maze lines. Each touch is counted as an error. Score is number of openings negotiated minus the number of errors. One 25-sec. trial.

35. *Steadiness.*—The *S* must trace between a pair of narrowly separated lines ($\frac{1}{16}$ in.) which form a pattern. Score is the number of segments negotiated without touching the lines. Two trials, no time limit.

36. *Discrimination Reaction Time—Printed.*—This is a printed version of the Discrimination Reaction Time Test (test 25 in this series). The *S* is provided with a series of items. Each item represents a stimulus setting. There are four possible correct responses to each setting. The *S* goes from item to item as rapidly as possible checking the appropriate response. Score is the number of items completed minus the number of errors. Two 60-sec. trials.

37. *Marking Accuracy.*—Described in (10). The *S* is given a standard IBM answer sheet in which one of the alternatives to each item has been overprinted with a small circle. The alternatives which are circled are randomly determined from item to item. The *S*'s task is merely to mark the answer sheet as rapidly as possible under the indicated circles. Score is the number of items completed minus errors. Two 40-sec. trials.

38. *Log-Book Accuracy.*—Described in (10). The *S* is provided with a test booklet. Items in the booklet merely pair an alternative letter with each item number. The *S* must find the number on a separate IBM answer sheet and mark in the indicated alternative to each item as rapidly as possible. Score is the number of correct items minus errors. Two 120-sec. trials.

TABLE 1

MEANS, *SD*'S, AND RELIABILITIES OF THE EXPERIMENTAL TESTS

Tests	Mean	*SD*	Reliability*
1. Precision-Steadiness			
Counter	370.2	183.1	.90
Clock	28.0	19.2	.85
2. Steadiness-Aiming	50.6	22.7	.84
3. Track-Tracing			
Counter	94.7	42.1	.91
Clock	7.8	3.9	.88
4. Two-Plate Tapping	717.0	83.7	.99
5. Key Tapping	1123.2	114.1	.93
6. Ten-Target Aiming			
Errors	81.3	41.6	.94
Corrects	364.5	43.7	.91
7. Rotary Aiming	447.7	60.5	.96
8. Hand-Precision Aiming			
Errors	158.5	101.7	.99
Corrects	337.1	54.1	.92
9. Visual Reaction Time	.246	.026	.86
10. Auditory Reaction Time	.233	.026	.85
11. Minnesota Rate of Manipulation (Placing)	98.5	7.0	.87
12. Minnesota Rate of Manipulation (Turning)	104.1	8.7	.79
13. Purdue Pegboard—Right Hand	16.9	1.9	.70
14. Purdue Pegboard—Left Hand	16.1	1.6	.68
15. Purdue Pegboard—Both Hands	26.0	3.0	.70
16. Purdue Pegboard—Assembly	37.3	5.5	.74
17. O'Connor Finger Dexterity	166.0	19.0	.76
18. Santa Ana Finger Dexterity	125.8	12.6	.91
19. Punch Board	178.7	20.8	.85
20. Pin Stick	75.5	11.5	.77
21. Dynamic Balance			
Matchings	17.4	13.0	.86
Time (sec.)	34.3	8.4	.82
22. Postural Discrimination—Vertical	23.6	10.5	.93
23. Postural Discrimination—Angular	39.0	14.2	.72
24. Medium Tapping	40.3	11.3	.91
25. Large Tapping	151.2	25.2	.94
26. Aiming	108.6	17.9	.89
27. Pursuit Aiming—I	69.5	12.6	.90
28. Pursuit Aiming—II	211.0	37.6	.93
29. Square Marking	139.3	30.6	.92
30. Tracing	47.2	11.6	.85
31. Steadiness	26.5	12.0	.84
32. Printed Discrimination Reaction Time	45.8	14.2	.87
33. Marking Accuracy	109.4	22.9	.91
34. Log-Book Accuracy	62.0	13.2	.79

* Reliability estimates are odd-even trial correlations corrected by the Spearman-Brown formula for full length of test, except where only one continuous trial is given. In these cases coefficients are test-retest coefficients.

PRETESTING THE TESTS

Procedure.—Each of the experimental apparatus tests was administered to separate groups of 200 basic trainee airmen. All the printed tests were administered to a single group of 200 basic airmen. The pretest analysis was aimed chiefly at determining the reliabilities of the tests, evaluating apparatus differences between models of the same test (since in all but a few cases, four models of each test were constructed for group testing), determining practice effects occurring within the testing period on each test, and verifying for each test the extent to which adequate differentiation of individual differences among *S*s is achieved.

Results.—The means, *SD*'s, and reliability estimates obtained are presented in Table 1. More extensive data on these tests have been presented in another report (2).

In general, it was found the tests possessed adequate reliabilities for inclusion in a factor analysis battery. Highest reliabilities were found for tests hypothesized to sample Tapping and Aiming areas, with lower reliabilities among the "dexterity" tests (due in some measure to the shorter testing periods). Reliabilities of the printed tests were quite high. Another finding was that in those tests yielding error counts along with some other score, the error count generally provided the more reliable measure of performance. The reliability estimates presented also are useful for comparison purposes with the test communalities obtained in the subsequent factor analysis.

The *SD*'s of the tests generally indicated that the measures are sensi-

TABLE 2

Test Intercorrelations and Final Residuals

TABLE 3

Centroid Factor Loadings

Variable	Factor I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	h^2
1. Steadiness-Precision	29	25	13	-19	-09	-16	-14	09	21	-10	-12	-13	34
2. Steadiness-Aiming CM105E	30	28	09	-20	-08	-22	-15	07	22	-16	14	09	40
3. Track-Tracing	37	26	15	-21	-16	-08	-21	05	13	-22	06	06	41
4. Two-Plate Tapping CM202A	61	-14	06	-22	-05	08	13	-25	07	19	-09	-06	58
5. Key Tapping	37	05	11	-13	-08	-05	17	-10	-10	12	-10	-22	29
6. Ten Target Aiming--errors	11	27	-21	25	-14	-28	23	12	11	12	23	-07	43
7. Ten Target Aiming--corrects	49	-26	27	-22	20	17	-06	-06	09	08	-24	18	49
8. Rotary Aiming	59	-28	09	-09	08	09	05	-13	09	05	13	10	51
9. Visual Reaction Time	44	-19	24	08	-27	32	-13	19	-23	18	13	-05	63
10. Auditory Reaction Time	30	-20	25	13	-27	15	-08	20	-25	23	14	-11	30
11. Minnesota Rate of Manipulation, Placing	64	-12	19	11	03	09	09	-07	15	04	08	-06	52
12. Minnesota Rate of Manipulation, Turning	67	02	09	17	09	02	14	06	02	09	-05	07	54
13. Purdue Pegboard, Right Hand	49	-10	33	28	-03	05	04	-07	06	-14	-04	-07	48
14. Purdue Pegboard, Left Hand	48	09	16	21	20	03	05	-13	-11	-22	03	06	43
15. Purdue Pegboard, Both Hands	54	10	26	31	16	05	-05	-09	-11	-09	08	08	54
16. Purdue Pegboard, Assembly	57	12	10	28	15	05	08	09	-12	-14	09	13	52
17. O'Connor Finger Dexterity	57	07	26	22	15	01	07	-07	-06	-10	-08	-05	49
18. Santa Ana Dexterity CM116A	48	-02	-06	08	22	-20	26	07	06	12	07	12	44
19. Hand Precision Aiming--errors	31	41	-21	29	-39	12	05	-33	25	11	-04	14	76
20. Hand Precision Aiming--corrects	26	-34	20	-29	33	-23	12	33	-12	-12	09	-11	64
21. Discrimination Reaction Time CP611D2	54	28	-24	-03	16	23	11	16	-07	-05	-08	-14	58
22. Complex Coordination CM701E	53	36	-11	-07	-04	17	03	09	-05	-13	-20	-09	54
23. Rudder Control CM120C	23	28	-16	-24	-13	-06	03	08	-20	09	-15	06	31
24. Rotary Pursuit CM603B	48	15	04	-19	-08	-07	10	03	-16	17	-21	22	46
25. Dynamic Balance	20	20	-04	-10	-04	19	-08	04	-09	08	-06	22	20
26. Postural Discrimination--Vertical	24	21	-18	-09	15	07	-19	-11	-14	11	17	-11	89
27. Postural Discrimination--Angular	15	23	-12	-17	-22	10	-23	-12	-03	18	22	-10	33
28. Punch Board	62	-08	15	02	-18	-07	-04	-09	15	02	-08	-13	50
29. Pin Stick	52	03	06	-03	10	-14	06	-11	04	08	-08	08	34
30. Medium Tapping	67	-25	-14	-18	-10	-11	24	-06	-09	-12	22	11	73
31. Large Tapping	64	-25	-14	-19	-11	-13	19	-13	-13	-14	16	04	67
32. Aiming	67	-40	-16	05	08	-16	-21	-16	-01	08	-12	06	76
33. Pursuit Aiming I	65	-40	-17	12	05	-23	-15	-15	06	05	-10	-05	74
34. Pursuit Aiming II	66	-35	-26	13	09	-22	-27	-07	-11	-04	-13	02	80
35. Square Marking	47	-20	-13	07	-13	05	-08	13	04	-11	-11	05	36
36. Tracing	45	-06	-15	-05	-15	16	-11	-09	11	-20	-03	04	33
37. Steadiness	34	10	06	-05	-18	-06	-18	02	10	10	10	-08	23
38. Printed Discrimination Reaction Time	50	19	-24	-03	07	08	15	19	17	05	-01	-13	46
39. Log-Book Accuracy	58	-12	-09	11	11	13	14	21	26	02	05	-06	62
40. Marking Accuracy	60	-21	-26	08	10	11	-08	11	12	-17	-08	-04	57
$\Sigma a^2/k$	25	05	03	03	03	02	02	02	02	02	02	01	

tive enough to discriminate differences among *Ss*. Moreover, individual differences continued to be demonstrated at various stages of practice on the tests.

In addition, scores were recorded by apparatus when there were several models of the same test. Differences between mean scores obtained as a function of apparatus model were tested by analysis of variance for each of these tests. In the case of only one test did the *F* ratio approach statistical significance. This was for the Steadiness-Aiming Test, where differences among the four test models were significant between the 5% and 1% levels of confidence.

Administration and Factor Analysis of the Complete Battery

Administration procedure.—The complete battery of 38 tests was administered to another sample of 400 basic trainee airmen. Two hundred of the *Ss* received the tests in one administration order, while the remaining 200 *Ss* received exactly the reverse administration order. This was done in order to reduce (as far as was feasible) possible influences due to the particular administration order of tests used.

Data analysis procedures.—The 38 tests in the battery yielded 40 separate scores. This was because two of the tests each yielded two scores (errors and corrects).[3] However, in the case of these tests, efforts were made to keep the measures experimentally independent. Thus, no two scores were ever taken off the same trial. For example, in the Ten-Target Aiming Test, *E* cumulated error scores for the odd trials and correct scores for the even trials. It was thought advisable to include both error and correct scores on these tests since it was suspected that each might possibly yield measures of somewhat different aspects of performance.

For each test the obtained distributions of raw scores were transformed to normalized distributions of standard scores (stanines), each with a range from 1 to 9, a mean of 5, and *SD* of 2. Conversions were made so that the nine end of the scale was always indicative of "good" performance (e.g., low errors, high time-on-target, low reaction time). Pearson product-moment correlations among the 40 variables were then obtained.

TABLE 4

Distribution of Twelfth Factor Residuals

Residual	Frequency
−.07	1
−.06	2
−.05	11
−.04	31
−.03	42
−.02	117
−.01	120
.00	165
.01	117
.02	106
.03	32
.04	25
.05	6
.06	3
.07	1
.08	1

The resulting correlation matrix is presented in the upper half of Table 2. Factors were extracted from the matrix by Thurstone's centroid method (18). Extractions were continued beyond the point where the product of the two highest centroid loadings became less than the standard error of the original correlation of the same two variables. Twelve centroid factors were extracted and the twelfth factor residuals computed. Table 3 presents the complete centroid factor matrix obtained. The bottom portion of Table 2 presents the final residual correlations. These residuals have been converted to a frequency distribution in Table 4. (Mean = −.001, *SD* = .021.) It is apparent that all the meaningful factor variance has been extracted from the original matrix.

Orthogonal rotations of the primary axes were accomplished using Zimmerman's graphical method (19). Since certain hypotheses regarding the possible factors were set up before the analysis, it was considered advisable in the present instance to preserve objectivity by completely "blind" rotations. Hence, the centroid matrix (minus variable names) was given to a skilled analyst[4] who had no knowledge whatsoever of the nature of the variables. Rotations were continued until the criteria of simple structure and positive manifold were closely approximated. The resulting clusters were then interpreted for psychological meaningfulness.

[3] Although the Precision-Steadiness and Track-Tracing Tests also yielded two scores (number of error contacts and time in contact), only one score, the more reliable one, was utilized from each of these tests.

[4] The writer is indebted to Mr. Gabriel Friedman for his part in this phase of the study.

TABLE 5

Rotated Factor Loadings

Variable	Factors* I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	h^2
	MPS	FD	RR	AI	St	MT	MD	PM	PC	SR	FD	D	
1. Steadiness-Precision	-02	01	11	09	**29**	-02	03	-02	07	25	02	04	14
2. Steadiness-Aiming CM10E	06	07	06	02	**60**	-05	03	11	07	02	15	01	41
3. Track-Tracing	19	21	21	05	**41**	12	-05	13	16	15	13	-02	46
4. Two-Plate Tapping CM20A	**36**	02	**29**	15	05	12	24	02	12	21	02	-10	39
5. Key Tapping	23	03	19	03	06	15	24	-16	12	28	-03	01	29
6. Ten Target Aiming--errors	06	-03	**-33**	-10	06	-22	**43**	15	01	07	19	-04	44
7. Ten Target Aiming--corrects	06	15	**66**	24	02	07	05	11	15	00	-04	19	60
8. Rotary Aiming	**36**	18	**46**	14	04	15	20	25	-03	-02	08	01	52
9. Visual Reaction Time	11	12	19	09	-02	**73**	02	15	09	05	02	02	61
10. Auditory Reaction Time	10	04	05	09	-03	**68**	11	02	02	00	-02	05	50
11. Minnesota Rate of Manipulation, Placing	22	**31**	**36**	15	12	22	**32**	20	-06	14	-01	-06	33
12. Minnesota Rate of Manipulation, Turning	20	**34**	21	19	09	17	**38**	24	21	15	-01	02	54
13. Purdue Pegboard, Right Hand	11	**46**	22	19	14	25	19	05	-09	10	-18	-05	48
14. Purdue Pegboard, Left Hand	19	**38**	13	08	06	02	13	08	08	08	07	02	44
15. Purdue Pegboard, Both Hands	10	**61**	14	15	09	19	21	06	10	02	12	-02	54
16. Purdue Pegboard, Assembly	17	**22**	03	07	08	15	21	25	18	07	06	07	32
17. O'Connor Finger Dexterity	13	**22**	19	16	10	13	25	02	10	18	-01	05	49
18. Santa Ana Dexterity CM116A	26	16	09	10	05	09	**47**	25	15	-01	05	11	44
19. Hand Precision Aiming--errors	02	13	01	02	17	01	15	13	26	13	00	-**77**	76
20. Hand Precision Aiming--corrects	27	03	14	04	09	04	16	07	-14	01	-01	**62**	63
21. Discrimination Reaction Time CP611E	16	21	05	01	00	02	10	**32**	28	**32**	05	18	50
22. Complex Coordination CM701E	12	25	09	03	19	05	-01	17	**36**	**31**	07	-05	54
23. Rudder Control CM120C	16	-11	-03	-02	14	02	-01	-02	**45**	20	02	10	31
24. Rotary Pursuit CM80B	21	04	22	08	16	10	17	05	**53**	04	04	00	45
25. Dynamic Balance	-01	05	13	-04	05	10	-08	13	**32**	05	-07	13	20
26. Postural Discrimination--Vertical	10	07	03	04	-01	03	02	00	08	18	**48**	-05	29
27. Postural Discrimination--Angular	-01	-02	11	-02	00	00	05	-01	05	12	**55**	-05	35
28. Punch Board	25	15	29	29	**39**	22	21	05	00	20	-07	-12	50
29. Pin Stick	23	19	25	19	**46**	-03	28	05	20	07	06	00	54
30. Medium Tapping	**74**	10	22	03	13	11	18	26	08	01	02	02	74
31. Large Tapping	**74**	10	21	08	12	09	11	16	05	06	05	03	68
32. Aiming	**45**	12	28	**63**	01	08	13	18	04	-02	11	00	76
33. Pursuit Aiming I	**29**	14	16	**63**	00	09	18	11	-01	02	05	00	73
34. Pursuit Aiming II	**48**	18	08	**60**	02	03	06	19	07	05	12	06	80
35. Square Marking	29	12	07	30	07	19	-05	**39**	11	12	-10	01	36
36. Tracing	28	14	21	16	15	07	-15	27	04	17	01	-18	35
37. Steadiness	07	-01	10	13	**31**	22	10	06	02	10	15	-11	24
38. Printed Discrimination Reaction Time	14	04	07	05	13	-02	26	**58**	15	**42**	11	-02	47
39. Log-Book Accuracy	23	08	12	20	00	05	28	**29**	-01	29	01	-01	43
40. Marking Accuracy	29	20	15	**37**	02	03	01	**46**	-01	26	02	04	37
$\Sigma a^2/r$	08	06	05	05	04	04	04	04	03	05	02	03	

* Factors are identified as follows: I. Wrist-Finger Speed; II. Finger Dexterity; III. Rate of Arm Movement; IV. Aiming; V. Steadiness; VI. Reaction Time; VII. Manual Dexterity; VIII. Psychomotor Speed; IX. Psychomotor Coordination; X. Spatial Relations; XI. Postural Discrimination; XII. Doublet.

The complete rotated factor matrix is presented in Table 5.

Interpretation of Factors

The 12 factors identified will be described in turn. Tests with loadings of .30 or higher are listed under each factor. Also presented are the loadings of each test (above .30) on other factors. A test with no loading listed on any other factors is a relatively pure measure of the factor discussed in the present analysis.

Factor I is obviously the same factor as that previously called Tapping or Speed of Simple Ballistic Movement. A better name would appear to be *Wrist-Finger Speed* (WFS), since all of the tasks representing this factor require rapid wrist flexing and finger movements. Either rapid rotary or bending movements of the wrist are required in all these tasks. Tests which do not emphasize such movements do not appear on this factor. This factor apparently represents simply the speed with which such rapid wrist-finger movements can be made. It appears to be best measured by printed tests which require *Ss* to tap rapidly with a pencil where careful positioning of the pencil is not required (e.g., tapping in larger circles). These results confirm earlier findings that this factor does not involve eye-hand coordination. Thus, loadings decrease systematically as the circles to be dotted in become smaller. The presence of this factor in certain of the apparatus tests suggests that tapping is too narrow a name for the factor which apparently is not so specific as was formerly suspected. Moreover, emphasis now appears centered on the wrist flexing movement rather than on simple speed of oscillation of the fingers or arm.

Factor I

Test No.	Test Name	Loading	Other Factors
31	Large Tapping	.74	—
30	Medium Tapping	.74	—
33	Pursuit Aiming—I	.50	.63 Ai
34	Pursuit Aiming—II	.48	.68 Ai
32	Aiming	.45	.63 Ai
4	Two-Plate Tapping	.36	.54 RM
8	Rotary Aiming	.36	.46 RM

Factor II is identified as *Finger Dexterity* (FD) or *Fine Dexterity*. It involves the ability to coordinate finger movements in performing fine manipulations. The tests most heavily loaded with this factor all require the grasping, releasing, or manipulation of small objects (e.g., pegs or pins) in which the finger tips are involved primarily. For example, in the Purdue subtests, the pegs and assembly components to be manipulated are quite small. In the O'Connor test, three small pins (and only three) must be picked up together from a tray of pins and manipulated properly.

It can also be noted that the first three variables on this factor and the Turning test all involve some use of the left hand, either alone or in combination with the right hand. Moreover, these are the only tests in the battery which require the use of the nonpreferred hand. This suggests that ambidexterity may be involved in this factor. The possibility also exists that this factor is a composite

Factor II

Test No.	Test Name	Loading	Other Factors
18	Purdue Pegboard, Both Hands	.61	—
14	Purdue Pegboard, Left Hand	.58	—
16	Purdue Pegboard, Assembly	.55	—
17	O'Connor Finger Dexterity	.53	—
13	Purdue Pegboard, Right Hand	.46	—
12	Minnesota Rate of Manipulation, Turning	.34	.38 MD
11	Minnesota Rate of Manipulation, Placing	.31	.32 MD, .36 RM

of fine dexterity and "a separate ambidexterity factor." Future analyses with additional reference variables should explore this possibility further.

Factor III is defined as *Rate of Arm Movement* (RM).

Factor III

Test No.	Test Name	Loading	Other Factors
7	Ten Target Aiming—corrects	.66	—
4	Two-Plate Tapping	.54	.36 WFS
8	Rotary Aiming	.46	.36 WFS
11	Minnesota Rate of Manipulation, Placing	.36	.31 FD, .32 MD
6	Ten-Target Aiming—errors	−.35	.43 MD

The factor appears confined to apparatus tests all of which emphasize the speed with which rather gross rapid arm movements can be made. Although all these tests were designed to measure certain other of the hypothesized factors, it seems evident in retrospect that the ability to make such rapid arm movements is the common feature of these tasks. In the Ten-Target Aiming and Rotary Aiming Tests, for example, the *Ss* must thrust quickly from one target to the next. The Two-Plate Tapping Test requires extensive arm movement. The finding that the Placing test has a much higher loading than the Turning test is also consistent with the present interpretation, since considerably more arm movement is required in the Placing test. Of special interest is the negative loading of the Ten Target Aiming error score. This suggests quite reasonably that rapid arm movements on the test are related to inaccuracy in hitting the successive targets.

Factor IV is identified only by certain printed tests which require marking with a pencil exactly in a specified small area. This factor appears to be the same as that called

Factor IV

Test No.	Test Name	Loading	Other Factors
34	Pursuit Aiming—II	.68	.48 WFS
33	Pursuit Aiming—I	.63	.50 WFS
32	Aiming	.63	.45 WFS
40	Marking Accuracy	.37	.46 PS
35	Square Marking	.30	.30 PS

Aiming (Ai) or Controlled Manual Movement in previous analyses. It has been defined as the ability to perform quickly *and* precisely a series of accurately directed movements requiring eye-hand coordination. It is distinguished from the wrist-finger speed factor in that this latter factor does not involve eye-hand coordination or visual alignment. Thus, loadings on the Aiming factor are highest in tests where the circles to be dotted become smaller, and insignificant in tests involving larger circles. The essential feature of this factor appears to be that both accuracy and speed are required in the restricted movement of the hand precisely from one place to another. However, there is no evidence to indicate that this factor extends beyond printed tests. Apparatus tests designed to measure this factor fail to show significant loadings on it. Instead, the variance in these apparatus tests is split upon other factors.

Factor V is clearly an *Arm-Hand Steadiness* (St) factor, and is defined by those tests designed to measure this hypothesized factor. It is best measured by tests which emphasize precise, accurate arm-hand movements of the type which minimize strength and speed. This factor appears common to such movements regardless of the plane of movement

Factor V

Test No.	Test Name	Loading
3	Track-Tracing	.61
2	Steadiness-Aiming	.60
1	Steadiness-Precision	.50
37	Steadiness (printed)	.31
28	Punch Board	.30

involved. For example, the Steadiness-Precision Test requires steady arm movements in a straight line to and away from the body, the Track-Tracing Test requires highly controlled arm movements in a lateral plane, and Steadiness-Aiming requires a more static kind of controlled steadiness (holding a stylus so as not to touch the sides of a small hole). Of further interest is the presence on this factor of the printed test specifically designed to measure steadiness, although the loading of this test is substantially lower (.31) than those of the apparatus steadiness tests.

Factor VI is confined only to the two reaction time tests and confirms earlier findings that *Reaction Time* (RT) is usefully considered a separate factor. It represents simply the speed with which an individual can react to a stimulus when it appears.

Factor VI

Test No.	Test Name	Loading
9	Visual Reaction Time	.73
10	Auditory Reaction Time	.68

Factor VII is identified as *Manual Dexterity* (MD). The tests loaded on this factor involve manipulating larger blocks or objects in which more of the whole hand is used to grasp, release, turn, or position the objects. The factor may be defined as the ability to make skillful arm-hand movements. It differs from Finger Dexterity in that fine manipulations with the finger tips are not so much involved. The skillful movements involved in this factor are more gross than those involved in Finger Dexterity. The separation of these two factors confirms the findings of previous studies.

Factor VII

Test No.	Test Name	Loading	Other Factors
18	Santa Ana Finger Dexterity	.47	—
6	Ten Target Aiming—errors	.43	−.35, RM
12	Minnesota Rate of Manipulation, Turning	.38	.34 FD
11	Minnesota Rate of Manipulation, Placing	.32	.31 FD, .36 RM

Factor VIII is best defined by two printed tests which emphasize simply speed of marking an answer sheet. These two tests have defined a factor tentatively labeled *Psychomotor Speed* (PS) in previous Air Force analyses. The present analysis does little to clarify the status of this factor beyond revealing its distinction from the other factors, especially from Aiming. The designation, Psychomotor Speed, seems too broad, however, since the factor is not very much involved in the apparatus psychomotor tests which emphasize speed. Whatever the precise nature of this factor, its significance is quite restricted.

Factor VIII

Test No.	Test Name	Loading	Other Factors
39	Log-Book Accuracy	.59	—
40	Marking Accuracy	.46	.37 Ai
38	Printed Discrimination Reaction Time	.38	.42 SR
21	Discrimination Reaction Time	.32	.53 SR
35	Square Marking	.30	.30 Ai

Factor IX is the factor previously called *Psychomotor Coordination* (PC). It consistently has been found best measured by the Rotary Pursuit Test and also by the Rudder Control and Complex Coordination Tests. The present results confirm this and also reveal that Dynamic Balance is loaded with this factor. The factor has been defined rather broadly as representing either coordination of the large muscles of the body, in move-

Factor IX

Test No.	Test Name	Loading	Other Factors
24	Rotary Pursuit	.53	—
23	Rudder Control	.45	—
22	Complex Coordination	.36	.51 SR
25	Dynamic Balance	.35	—

ments of moderate scope, or coordination of such movements with the perception of a visual stimulus. It was hoped that the inclusion of the wide range of psychomotor tests in the present study would throw additional light on the nature of this factor. It was hypothesized, for example, that the factor of Psychomotor Coordination might be split into several more basic factors. The present analysis has failed to effect this split, however, and indicates that this factor is quite independent of the other psychomotor factors identified.

Factor X is defined by the Discrimination Reaction Time and Complex Coordination Tests. These tests have consistently appeared together on a *Spatial Relations* (SR) factor. The essential feature of this factor appears to be the ability to relate different responses to different stimuli where either stimuli or responses are arranged in spatial order. Of special interest is the presence of the printed version of the Discrimination Reaction Time Test on this factor. Emphasis in this Spatial factor appears to be on decision as to direction of movement.

Factor X

Test No.	Test Name	Loading	Other Factors
21	Discrimination Reaction Time	.53	.32 PS
22	Complex Coordination	.51	.36 PC
38	Printed Discrimination Reaction Time	.42	.38 PS

Factor XI is a doublet confined to the two tests given on the Postural Discrimination apparatus. For the present it may be called *Postural Discrimination* (PD) and defined as the ability to make use of postural cues in

Factor XI

Test No.	Test Name	Loading
27	Postural Discrimination—Angular	.55
26	Postural Discrimination—Vertical	.48

making precise bodily adjustments in the absence of visual cues. This is similar to the definition given earlier of the hypothesized motor-kinesthesis factor. This factor was not found in two other tests (Rudder Control and Dynamic Balance) where its presence in some degree was suspected. Both these latter tests allow the use of visual cues, however, while in the two Postural Discrimination Tests the *S* wears dark goggles. Whether this factor extends beyond gross bodily adjustments to accurate adjustments of the limbs, for example, remains to be discovered.

Factor XII is a doublet factor confined only to the error and correct scores on the Hand-Precision Aiming Test. Sources of variance in these variables have thus been obscured by the inclusion of both scores in the present analysis. Therefore, no further significance is attributed to this factor.

Factor XII

Test No.	Test Name	Loading
19	Hand-Precision Aiming—errors	−.77
20	Hand-Precision Aiming—corrects	.69

Discussion

In general, the results of this study have verified in a single analysis certain factors hypothesized from the results of several previous studies. There were, however, some exceptions. For example, the factor tentatively called Psychomotor Precision in certain previous Air Force analyses failed to appear. This factor had been defined by the Discrimination Reaction Time and Santa Ana Finger Dexterity Tests and sometimes also by the Rotary Pursuit Test. It is possible that the variance in these tests due to the factor previously called Psychomotor Precision has been split up more meaningfully on certain of the other factors identified in the present analysis. In addition, one new factor, Rate of Arm Movement, was identified to account for

certain of the relationships.[5] The appearance of a Spatial Relations factor, while not listed as one of the hypothesized factors, is not surprising since this factor has consistently appeared in the two tests which define it in the present analysis. It should also be pointed out that the factor identified as Postural Discrimination may not be the same as the factor of Motor-Kinesthesis previously hypothesized, since in the present study it was confined to two tests given on a single apparatus. Its significance beyond these tests needs to be determined.

With respect to the individual tests in the battery, it may be said that they generally turned out to sample the factors they were designed to measure. For example, certain of the Arm-Hand Steadiness tests were practically pure measures of this factor. Similar successes were achieved in other areas. Most notable failures were the apparatus tests designed to measure Aiming. Instead, variance in these tests was without exception better accounted for by other of the factors identified.

In general, the communalities for the various individual tests are not high. In the case of certain tests, over half the variance remains unaccounted for by the factors identified in the present study. While some of this variance is error variance, the reliabilities of these tests are generally quite high, which indicates either that these tests contain a large amount of specific variance or that additional common factors remain to be identified in these tests.

Another objective of the study was to determine the utility of certain specifically designed printed tests in reproducing factor variance found in their apparatus test counterparts. The results show that the printed tests of the Wrist-Finger Speed factor are better measures of this factor than are the apparatus tests loaded on this factor. Similarly, loadings of any appreciable size on the Aiming factor are found only among certain printed tests. The attempt to design a printed test of Steadiness was moderately successful, although the loading of the test falls short of that achieved by the apparatus tests of this factor. However, this printed test of Steadiness might be substituted, where it is administratively unfeasible to use the apparatus tests. A similar conclusion might be reached in the case of the printed version of the apparatus Discrimination Reaction Time Test, since both these tests apparently possess about the same factor structure. However, the apparatus version is still the better measure of Spatial Relations, the main factor held in common by these two tests.

The study also sought to test the possibility that the factor called Psychomotor Coordination might be broken down into more basic ability categories. This factor had consistently been found in certain Air Force psychomotor tests (Rotary Pursuit, Rudder Control, and Complex Coordination). Previous analyses have not allowed for this possibility owing to the limited range of psychomotor tests employed in these batteries. As indicated earlier, the objective of splitting up this factor was not achieved in the present study.[6] Even with the wide range of psychomotor tests included, Psychomotor Coordination appeared again as a separate and quite independent factor. Moreover, none of the other factors identified in the present study was found to contribute significant variance to performance on these more complex-appearing psychomotor tasks.

Summary

The present paper has described a large-scale factor analysis study of performance in the relatively unexplored aptitude area of psychomotor skill. A large number of apparatus and printed psychomotor tests was selected or specifically designed to measure certain ability categories hypothesized to exist from the results of previous research in this area. After extensive pretesting, the tests were assembled into

[5] The importance of this factor has subsequently been verified in later analyses (6, 8).

[6] This objective has apparently been achieved in a later study (9), which included combinations of psychomotor variables quite different from those in the present study.

a battery and administered to 400 Ss. The correlations among scores on these tests were then subjected to a Thurstone centroid factor analysis.

The results of the study appear to confirm the existence of the following relatively independent factors in psychomotor skill:

(1) *Wrist-Finger Speed;* (2) *Finger Dexterity;* (3) *Rate of Arm Movement;* (4) *Manual Dexterity;* (5) *Arm-Hand Steadiness;* (6) *Reaction Time;* (7) *Aiming* (eye-hand coordination); (8) *Psychomotor Coordination;* (9) *Postural Discrimination;* and (10) *Spatial Relations.* Two additional factors were isolated but their interpretation or significance remains doubtful.

The results were discussed with respect to the factor composition of individual tests, the utility of certain printed tests designed to reproduce apparatus test variances, and the contribution of these factors to certain kinds of more complex psychomotor performance.

References

1. Fleishman, E. A. A factor analysis of intra-task performance on two psychomotor tests. *Psychometrika*, 1953, **18**, 45–55.
2. Fleishman, E. A. A factorial study of psychomotor abilities. *USAF, Personnel & Training Res. Cent., Res. Bull.*, 1954, No. 54-15. (AFPTRC-TR 54-15.)
3. Fleishman, E. A. An evaluation of two psychomotor tests for the prediction of success in basic flying training. *USAF, Hum. Resour. Res. Cent., Res. Bull.*, 1953, No. 53-9.
4. Fleishman, E. A. Testing for psychomotor abilities by means of apparatus tests. *Psychol. Bull.*, 1953, **50**, 241–268.
5. Fleishman, E. A. The perception of body position in the absence of visual cues. *J. exp. Psychol.*, 1953, **46**, 261–270.
6. Fleishman, E. A. The relation between abilities and improvement with practice in complex psychomotor performance. *J. exp. Psychol.*, in press.
7. Fleishman, E. A., & Hempel, W. E., Jr. A factor analysis of dexterity tests. *Personnel Psychol.*, 1954, **7**, 15–32.
8. Fleishman, E. A., & Hempel, W. E., Jr. Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, 1954, **19**, 239–252.
9. Fleishman, E. A., & Hempel, W. E., Jr. Factorial analysis of complex psychomotor performance. *USAF, Personnel & Training Res. Cent., Res. Bull.*, 1954, No. 54-12. (AFPTRC-TR 54-12.)
10. Guilford, J. P., & Lacey, J. I. (Eds.) *Printed classification tests.* AAF Aviation Psychology Program Research Report No. 5. Washington: Govt. Print. Off., 1947.
11. Hempel, W. E., Jr., & Fleishman, E. A. A factor analysis of physical proficiency and manipulative skill. *J. appl. Psychol.*, 1955, **39**, in press.
12. Hines, M., & O'Connor, J. A measure of finger dexterity. *J. pers. Res.*, 1926, **4**, 379–382.
13. MacQuarrie, T. W. A mechanical ability test. *J. pers. Res.*, 1927, **5**, 329–337.
14. Melton, A. W. (Ed.) *Apparatus tests.* AAF Aviation Psychology Program Research Report No. 4. Washington: Govt. Print. Off., 1947.
15. Minnesota Rate of Manipulation Test. Educational Test Bureau, 720 Washington Avenue S. E., Minneapolis, Minnesota.
16. Purdue Pegboard. Science Research Associates, 57 West Grand Avenue, Chicago, Illinois.
17. Reynolds, B. Correlations between two psychomotor tasks as a function of distribution of practice on the first. *J. exp. Psychol.*, 1952, **43**, 341–348.
18. Thurstone, L. L. *Multiple-factor analysis.* Chicago: Univ. Chicago Press, 1947.
19. Zimmerman, W. S. A simple graphical method of orthogonal rotation of axes. *Psychometrika*, 1946, **11**, 51–55.

(Received March 31, 1954)

Journal of Experimental Psychology
Vol. 48, No. 6, 1954

SOME EFFECTS OF WORK LOADING IN A REPETITIVE MOTOR TASK[1]

INA McD. BILODEAU AND EDWARD A. BILODEAU

Skill Components Research Laboratory, Air Force Personnel and Training Research Center

A clear inference from Hull's formulation of work inhibition (I_R) is that, response time and number of responses constant, the greater the work loading (W) of a response, the greater the inhibition accrued (6). In a self-paced task, however, the number of responses per unit time (or the time required to make a constant number of responses) is not constant across work loadings. The question may be raised whether differences in the performance of different work-loading groups reflect differential I_R; another possibility is that the decreased output with increased work loading compensates for the greater increment in I_R per response.

One approach to the question posed above is to explore the amount recovered over rest as a function of work loading (1, 2, 5). This approach is based on the expectation that the greater the inhibition accrued, the greater is the amount dissipated in a constant period of rest. The first pair of studies to be reported deals with the question of recovery in rate of responding on a cranking task, work-loading variable; the same loading was used in pre- and postrest practice. In the first study, the effect of loading was studied for five durations of a single rest, duration of prerest practice constant. In the second study, the recoveries of two work-loading groups were compared after five durations of prerest practice and two durations of rest, for four successive rests. The studies were designed to form, in combination, a fairly comprehensive survey of the effects of the usual distribution of practice variables on recovery under different work loadings. The first study employed a wide range of loadings and durations of rest, and a constant, extended duration of prerest practice; rest was introduced at a point where between-loading differences were fairly stable. The second study covered a number of brief durations of prerest practice, in the range where between-loading differences change most rapidly, and repeated the work-rest cycle to explore successive recoveries.

In the second pair of studies, a different approach was taken, that of comparing performances against one work loading for groups who had previously worked against different loadings. The expectation was that if no difference in inhibition develops during practice and habit strength is constant, then later performance, tested against a constant loading, would be independent of the prior loading. Previous studies of manual cranking have suggested that after instantaneous shifts in work loading, and shifts after brief rests, little or no difference in postshift performance could be attributed to prior work loading (2, 4). There was, however, some indication that the dependence of performance on prior work loading increased with a longer period of preshift practice and a longer rest before the shift in work loading; after 5 min.

[1] The opinions or conclusions contained in the present report are those of the authors. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

of cranking and a 4-min. rest, *Ss* switched from a heavy to a light loading cranked less rapidly than *Ss* practicing against the light loading both before and after rest (4).

The first of the two loading-shift studies to be reported used a relatively long preshift practice period of constant duration and a wider range of interpolated rests; the interpolated rests were concentrated around the duration yielding maximal between-loading differences in the studies mentioned above. In the second study, instantaneous shifts in work loading were performed after varying durations of preshift cranking; the study varied systematically the duration of preshift practice variable, with the shifts concentrated most heavily in the range where between-loading differences increase most markedly. Both investigations were restricted to a single shift from a heavy to a light work loading, with control groups working against the lighter loading throughout. The studies were designed to investigate performance after a shift from a heavy to a light loading as functions of the durations of (*a*) intershift rest and (*b*) preshift practice.

The simple, cranking response used in all four studies was selected on the assumption that the response was already well established and, therefore, that learning increments would have little effect on performance.

Method

Apparatus.—The apparatus was a manual crank, described in detail in previous studies (2, 3). Variation in work loading was effected by short-circuiting a tachometer generator which exerted a braking force on the crank shaft. Five work loadings—named Loads 0, 1, 2, 3, and 4, from least to greatest—were available. Rate of responding constant for a pair of loads, the ratio of their horsepower requirements was constant over all rates of rotation. Table 1 gives the horsepower required in Exp. I to turn the crank at several rates of rotation. All five loadings were employed in Exp. I; only Loads 0 and 4, the loading extremes, were used in Exp. II, III, and IV. Substitution of new gears and minor changes in the alignment of parts after Exp. I slightly changed the horsepower requirements, but left the apparatus otherwise unchanged. At 70 and 40 rev. per 20 sec. in Exp. II, III, and IV, for example, the horsepower requirements of Load 0 became .0286 and .0122, while for Load 4 the requirements became .0745 and .0300.

TABLE 1

Horsepower Required to Turn the Crank at Specified Rates of Revolution Against Five Different Work Loadings

Load	Revolutions per 20 Sec.				
	40	50	60	70	80
0	.0095	.0135	.0180	.0225	.0280
1	.0175	.0260	.0355	.0450	.0565
2	.0210	.0310	.0415	.0525	.0660
3	.0240	.0335	.0475	.0600	.0760
4	.0285	.0410	.0545	.0690	.0850

Regardless of loading, *S*'s task was to turn the crank with the preferred hand, under instructions to crank as fast as possible. To allow for inertia and *S*'s reaction time, *E* maintained a delay of about ½ sec. after giving the starting signal before starting the test clock. The number of revolutions completed in successive scoring periods was recorded from counters on the control panel. Two counters were provided to operate in alternate scoring periods.

Instructions.—The *Ss* were instructed and tested individually. The instructions were to crank as fast as possible between starting and stop signals. A ready signal was given about 5 sec. before the starting signal. The *Ss* were told neither the durations nor number of their practice and rest periods, nor were they given any information about work loading or changes in work loading.

Subjects.—The *Ss* were basic trainee airmen at Lackland Air Force Base. An *S* was used in only one experiment. In each experiment the *Ss* were assigned to a group according to successive, different testing sequences employing each group once (the control group of Exp. IV twice). A total of 1,690 *Ss* was used.

Procedures.—The designs and procedures of the four experiments are summarized in the following paragraphs.

In Exp. I, a factorial design, with five degrees of each of two variables, was used. The variables were (*a*) work loading (Loads 0,

1, 2, 3, and 4) and (*b*) the duration of rest (5, 10, 20, 40, and 80 sec.) interpolated between pre- and postrest practice periods. A different group of ten *S*s served in each of the 25 factorial combinations (groups). The experimental schedule involved: (*a*) a prerest period of 4 min. of continuous cranking against one of the five work loadings; (*b*) an interpolated rest of variable duration; and (*c*) a postrest period of 3 min. of continuous cranking against the loading prevailing in prerest practice. The number of revolutions completed in successive 20-sec. scoring periods (rev./20 sec.) was recorded.

In Exp. II, the design was a three-variable factorial, with 20 groups of 20 *S*s each. The variables were: (*a*) loading (Loads 0 and 4); (*b*) trial duration (10, 20, 30, 40, and 70 sec.); and (*c*) intertrial rest duration (5 and 40 sec.). The *S* received five trials (four intertrial rests), the same loading, trial duration, and rest duration obtaining for his five trials. The number of revolutions per 10 sec. was recorded.

In Exp. III, a two-variable, ten-cell, factorial design was used, in which the variables were (*a*) the work loading in preshift practice (Loads 0 and 4), and (*b*) the duration of rest interpolated between pre- and postshift practice (0, 1, 3, 6, and 60 min.). Forty *S*s were assigned to a group (cell). All *S*s received: (*a*) 5 min. of continuous preshift cranking (Load 0 or Load 4); (*b*) an interpolated rest of varying duration (including 0 min. rest); and (*c*) 2 min. of continuous postshift cranking against Load 0. Only half the *S*s, thus, received a shift in loading, although pre- and postshift practice are the terms used to differentiate the two periods for all groups. The number of revolutions per 5 sec. was recorded for the first minute of both pre- and postshift periods; and rev./10 sec. were taken after the first minute of practice. The *S*s with 1, 3, or 6 min. rest remained in the experimental room during the rest, while *S*s with a 60-min. rest were sent to a waiting room.

In Exp. IV, the single variable, differentiating seven groups, was the duration of practice against Load 4 preceding an instantaneous shift to Load 0. All *S*s received a single period of 4 min. of continuous cranking. A control group of 160 *S*s worked against Load 0 for the entire 4 min., while experimental groups of 80 *S*s each were shifted instantaneously from Load 4 to Load 0 after 5, 10, 15, 30, 60, or 120 sec. of cranking. After the shift from Load 4 to Load 0, *S* continued to work against Load 0 for the remainder of the 4-min. period. The

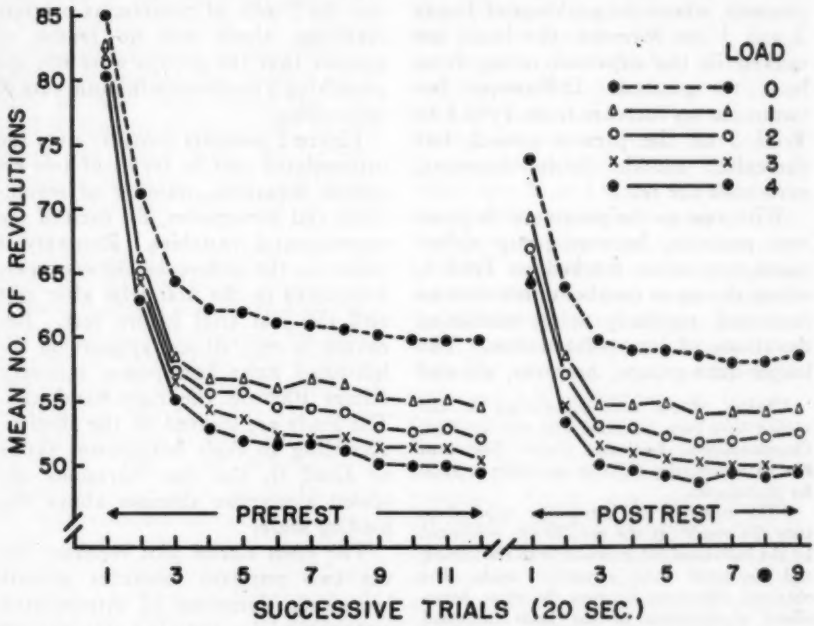


FIG. 1. Mean rev./20 sec. in pre- and postrest practice, with load as the parameter. The means have been averaged over five durations of interpolated rest.

number of revolutions per 5 sec. was recorded throughout the session.

Results[2]

Experiment I.—The following presentation does not deal with the individual conditions nor with all possible ways of expressing the data, but deals with selected main effects of work loading and rest. When one of the variables is considered, the data presented have been averaged over all values of the other variable. The means are therefore based on 50 cases —ten for each of the five values of the second variable.

In Fig. 1, mean rev./20 sec. is plotted against successive trials of the pre- and postrest periods. The effortfulness of the cranking task is evident in the negative slopes of the five curves. Between-loading differences in rate of cranking are apparent from the start; except on Trial 1 of prerest practice, where the rankings of Loads 2 and 3 are reversed, the loads are ranked in the expected order, from least to greatest. Differences between curves increase from Trial 1 to Trial 3 of the prerest period, but thereafter remain fairly constant, even over the rest.[3]

With rest as the parameter in postrest practice, between-group differences were most marked on Trial 1, where the mean number of revolutions increased regularly with increasing durations of interpolated rest. The longer rest groups, however, showed faster decrement than the groups with brief rests, and in the terminal trials the groups with 5-, 10-, and 20-sec. rests were responding at the same level. Although the longest rest group (80 sec.) remained slightly superior to the other groups throughout the 2 min. of continuous postrest cranking, there was no reason to suspect that the groups were not approaching a common minimum rate of responding.

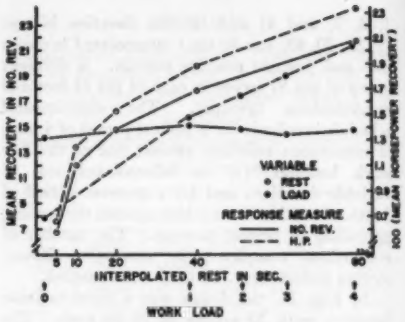


Fig. 2. Mean difference in rev./20 sec. and in horsepower requirements between the first prerest and the last postrest trials, as functions of work loading and duration of interpolated rest

Figure 2 presents recovery over the interpolated rest in terms of two response measures, number of revolutions and horsepower, for each of the experimental variables. Recovery is taken as the difference between performances in the first trial after rest and the last trial before rest. Recovery in rev./20 sec. appears on the left-hand axis; horsepower recovery (times 100) on the right-hand axis. The loads are located on the abscissa according to their horsepower ratios to Load 0; the rest durations are scaled along the abscissa above the loading scale.

The open circles plot recovery for the two response measures against increasing durations of interpolated rest: the solid curve presents recovery in rev./20 sec., the broken curve

[2] Tables of the data summarized in this section have been deposited with the American Documentation Institute. Order Document No. 4416, remitting $1.75 for microfilm or $2.50 for photocopies.

[3] The mean of the horsepower required to turn the crank (at the rev./20 sec. evidenced by the individual *S*s) increased with the loading and decreased over successive trials. The obtained differences between the mean horsepower requirements of the loads decreased rapidly in the early trials and less rapidly thereafter.

presents the horsepower measure. For either measure, recovery increases as a function of duration of interpolated rest. Both rest curves are similar in trend to the negatively accelerated, increasing rest functions obtained in learning studies for relatively brief durations of interpolated rest.

The trend of recovery with increasing work loading (closed circle) is not so simple, the two recovery measures showing different trends. The horsepower recovery function (broken curve) is of the same general form as the rest curves: the horsepower gain over the rest increases regularly as the work loading increases. Recovery in rev./20 sec. (solid lines), however, shows no trend. About 14 rev. are gained over the rest, regardless of loading. The lack of difference in recovery for different loads does not seem to be an artifact of an interaction of work loading and rest. A two-way analysis of variance of the gains in number of revolutions yielded an *F* ratio of less than unity for the Load × Rest interaction term, as well as for the Load main effect; rest duration, however, was significant well beyond the 1% level of confidence.

Experiment II.—In Exp. II the major variable for which recovery for different loadings was studied was the duration of prerest cranking, for durations much briefer than the duration used in Exp. I; only two durations of interpolated rest and only the work-loading extremes, Loads 0 and 4, were employed. The brief durations of prerest practice were chosen to test the independence of recovery (over loads) during the period for which there is a marked interaction of loads and trials, i.e., when the between-loads performance difference is changing most rapidly.

Figure 3 presents mean gain in

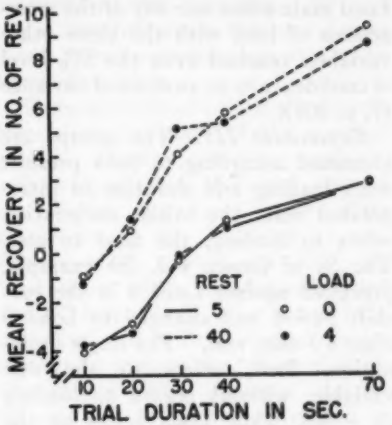


FIG. 3. Mean difference in number of revolutions between the first 10 sec. of Trial 2 and the last 10 sec. of Trial 1 as a function of trial duration, with intertrial rest and work loading as the parameters

number of revolutions between the last 10-sec. scoring period of Trial 1 and the first 10-sec. of Trial 2. Trial duration is the variable along the abscissa; load and duration of interpolated rest are the parameters. Except for the distortion introduced by the groups with 10-sec. trials, for which the Trial 1–Trial 2 differences are plotted, the four recovery functions are of the expected form: as duration of prerest practice increases, the amount recovered is greater and the difference between 5- and 40-sec. rest groups increases. Recoveries over the second, third, and fourth rests are not presented, as the between-load differences in recovery shown in Fig. 3 are representative of the differences in the remaining recoveries. There was no apparent difference between gains for Loads 0 and 4 for either experimental variable or for the different trials. The grand mean number of revolutions gained (averaged over all trials and all trial and rest durations) was 3.10 for Load 0 and 3.03 for Load 4. Neither the

Load main effect nor any of the interactions of load with the three other variables reached even the 5% level of confidence in an analysis of variance (7, p. 305).

Experiment III.—The groups are identified according to both preshift work loading and duration of interpolated rest; the initial designation refers to loading, the final to rest. The *S*s of Group 4–3, for example, practiced against Load 4 in the preshift period and changed to Load 0 after a 3-min. rest. The single designation, Rest, refers to the rest variable, without regard to loading ($n = 80$), while Load refers to the preshift loading variable, averaged over all values of the rest variate ($n = 200$).

Only the data of the postshift period, when all *S*s practiced against Load 0, will be treated in the following presentation. The preshift curves were of the same general form as the prerest curves presented in Fig. 1, although for the 5-sec. trials there was some slight positive acceleration in the decrement between Trials 1 and 3. Of interest was the finding that a difference between the grand means

Fig. 4. Mean rev./5 sec. in successive trials of the postshift period, with preshift work loading as the parameter. The means have been averaged without regard to duration of interpolated rest.

of Loads 0 and 4 was present even in the first 5-sec. trial, although (on Trial 1, only) the best group practicing against Load 4 exceeded the rate of the poorest group practicing against Load 0. The approximate between-load differences of 1.5 rev./5 sec. in the terminal trials of preshift practice and .5 rev. on postshift Trial 1 are a rough indication of the relative effects of present and past work loading.

The mean numbers of revolutions in successive 5-sec. trials of postshift practice are shown in Fig. 4. Five-second scores were recorded for Minute 1 (Trials 1–12), 10-sec. scores thereafter. In Fig. 4, the 10-sec. scores have been halved and are plotted as the means of successive pairs of trials (at points on the abscissa midway between the two trials). The open circles represent the grand means of the five Load 0 groups, the closed circles the grand means of the Load 4 (shift) groups. There is very little absolute difference between load grand means on any trial of postshift practice. The difference is greatest on Trial 1, but it is noteworthy that the grand mean of the Load 4 groups actually rises from the first to the second postshift trial. Nonetheless the mean of Load 0 exceeds the mean of Load 4 on all but two trials, though the difference is largely attributable to a single Load 4 group (4–6). Group 4–6 was considerably out of line with respect to other Load 4 groups, and performed at a lower rate than did Group 4–3 for about the first half of the postshift period.

In Fig. 5, rate of responding in the postshift period is plotted against duration of interpolated rest; preshift loading and ordinal number of the trial of postshift practice are the parameters. Each point represents the mean of a single group ($n = 40$).

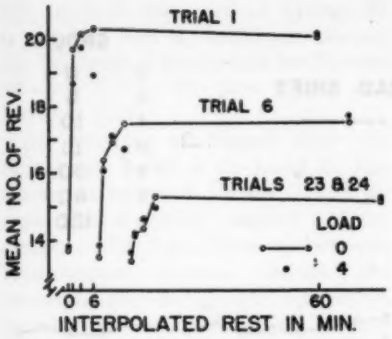


FIG. 5. Mean rev./5 sec. in postshift practice as a function of the duration of the interpolated rest, with work loading and ordinal number of the postshift trial as the parameters

To increase legibility, succeeding trials are displaced along the abscissa. The means of Load 0 groups are plotted as open circles, the means of groups shifted from Load 4 to Load 0 are plotted as unconnected, closed circles.

On Trial 1, the curve for Load 0 rises quickly, exhibiting the expected, negatively accelerated form; Group 0–60 has a slightly lower mean than does Group 0–6, but, otherwise, output immediately after rest is greater, the longer the rest. The rest trend is maintained in the later trials, though considerably reduced in magnitude. From trial to trial, the outputs of the Load 4 groups show trends similar to the trends of the Load 0 groups, except for the reversal of Groups 4–3 and 4–6, with most of the points falling fairly close to the corresponding points of the Load 0 curve.

Load comparisons on Trial 1 at the several values of rest show Groups 4–0 and 4–60 slightly superior to their respective controls, Groups 0–0 and 0–60, while Groups 4–1, 4–3, and 4–6 fall short of Groups 0–1, 0–3, and 0–6. There is, thus, no consistent trend toward increasing or decreasing differences between loads as the duration of interpolated rest increases. By and large, the differences between Loads 0 and 4, rest constant, are smaller in succeeding trials.

Three analyses of variance were performed on the data of the postrest period. The first, a two-way analysis, using only the data of Trial 1, showed the F for the Load main effect significant at the 5% level of confidence (4.03, $df = 1$ and 390), the Rest main effect significant beyond the 1% level (84.26, $df = 4$ and 390), and the Load × Rest interaction not significant at even the 5% level (1.35, $df = 4$ and 390). Separate analyses (7, pp. 281–284) were performed on the data of the first minute (Trials 1 through 12) and second minute of the postshift period. In both analyses, the F for the Rest main effect, with $df = 4$ and 390, was significant well beyond the 1% level of confidence ($F = 80.71$ for Minute 1 and 28.69 for Minute 2), as was the Rest × Trial interaction ($F = 28.96$, $df = 44$ and 4290, for Minute 1 and $F = 2.16$, $df = 20$ and 1950, for Minute 2). The F for the Load main effect, with $df = 1$ and 390, did not reach the 5% level in either analysis. In the analysis of the first minute of postrest practice, the F for the Load × Trial interaction (1.85, $df = 11$ and 4290) was significant at the 5% level of confidence, but no other interaction term involving load reached even the 5% level in the analysis of either minute. Statistically, therefore, there is little ground for rejecting the null hypothesis for load effects, while rest effects are highly significant.

Experiment IV.—The six experimental groups are identified as Groups 5, 10, 15, 30, 60, and 120 in terms of the duration of cranking (in seconds) preceding the instantaneous shift from Load 4 to Load 0. The control group, practicing against Load 0 throughout

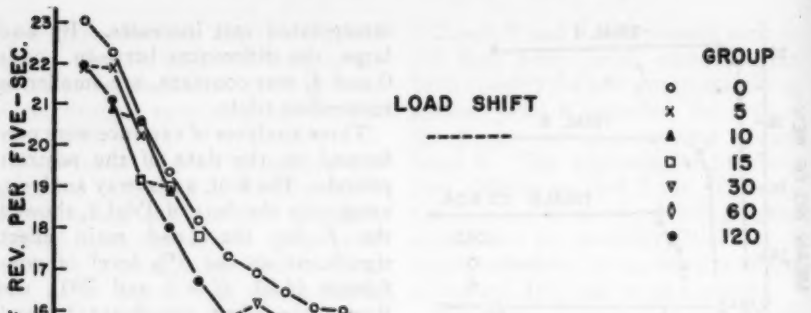


FIG. 6. Mean rev./5 sec. in the first 26 trials of continuous cranking, with duration of preshift practice against Load 4 as the parameter

the 4 min. of continuous practice, is called Group 0.

In Fig. 6 the mean rev./5 sec. are plotted for the first 130 sec. of continuous cranking for Groups 0 and 120. Three trials, only, are represented for the other groups: the last trial before the shift to Load 0, and the first two trials after the shift. The dashed line between points indicates the shift, in the case of each experimental group. The shift from Load 4 to Load 0, after any duration of preshift cranking, is accompanied by a decrease in the difference between the experimental group and Group 0; Groups 5, 10, and 15 show a slower rate of decrement over the shift, while Groups 30, 60, and 120 show an increment in rate of responding between the last preshift and the first postshift trials.

If the differences between the control group and the several experimental groups on the respective first postshift trials of the latter are used as indices of the amount of differential decrement to be attributed to the previous work loading, no clear-cut trend is evident. In terms of increasing differences from Group 0 on their respective first postshift trials, the experimental groups are ranked in the order: 10, 15, 5, 60, 30, and 120. No function, thus, can be specified, although the three groups shifted after working against Load 4 for no more than 15 sec. are less different from Group 0 on the first postshift trial than are the three groups shifted after cranking for 30 sec. or longer against Load 4. The dichotomy of relatively brief vs. relatively long durations of preshift work is borne out by t tests of the significance of the differences between the control group and the experimental groups on postshift Trial 1. The control vs. experimental group difference ($df = 238$) does not reach even the 5% level of confidence in the cases of Groups 5 ($t = 1.87$) and 10 ($t < 1.00$), is significant at the

5% level in the case of Group 15 ($t = 2.01$), and is significant beyond the 1% level of confidence for Groups 30 ($t = 3.71$), 60 ($t = 3.20$), and 120 ($t = 3.24$).

As cranking continued after the shift from Load 4 to Load 0, the difference between Group 0 and an experimental group tended to decrease, as did the differences between experimental groups, though the trends in between-groups differences were not smooth and regular. Nonetheless, except in the case of Group 5, there were few postshift trials in which an experimental group exceeded Group 0 in mean rate of responding.

The effect of duration of preshift cranking against Load 4 is illustrated in Fig. 7 where the mean difference between Group 0 and each experimental group is plotted against duration of preshift practice. The mean differences are for 1-min. intervals, i.e., the sum of the differences for 12 5-sec. trials; the open circles represent the first, the closed circles the second, minute of postshift practice. Time is thus held constant between experimental groups in terms of duration of postshift practice, e.g., above the 5-sec. duration on the abscissa, the Minute 1 point represents the mean difference between Groups 0 and 5 on Trials 2–13, while above the 120-sec. duration the point for Minute 1 represents the difference between Groups 0 and 120 on Trials 25–36. All differences in Fig. 7 are positive, indicating a greater output for Group 0 than for any experimental group. In the first minute of postshift practice the difference between the means of Group 0 and the experimental groups is, in general, an increasing function of duration of preshift cranking, though Group 60 is more unlike Group 0 than is Group 120. The differences decrease, by and large, from the first to the second minute, and though the trend is still roughly maintained, the ranking is less consistent for Minute 2 than for Minute 1.

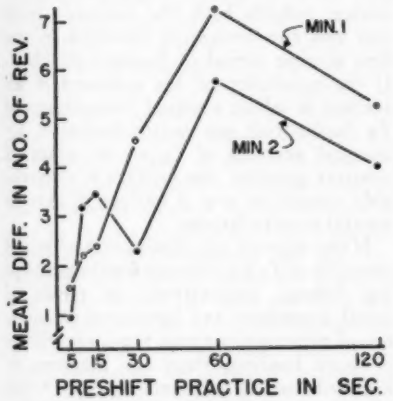


FIG. 7. Mean difference, control minus experimental, in revolutions per minute as a function of duration of preshift practice. The parameter is the ordinal number of the minute of postshift practice taken from different origins corresponding to the shifts from Load 4 to Load 0 for the different preshift durations.

If the groups are compared on the same trials (total duration of cranking against either load constant), Groups 0 and 5 fall into one rough grouping from Trial 6 on, Groups 10, 15, and 30 form another rough grouping after about the first minute of practice and look very little different from Group 0 in the last minute, or so, while Groups 60 and 120 remain discrete from Group 0 through the final minute. In terms of increasing mean number of revolutions in the fourth minute of practice, the groups are ranked in the order 120, 60, 10, 30, 15, 5, and 0. The ordering suggests that performance depends somewhat on the time elapsing since the shift from Load 4 to Load 0. On individual trials, however, there is considerable between-group confusion in postshift perform-

ance, whether time is held constant from the start of practice or from the point of the shift from Load 4 to Load 0. A Groups × Trials analysis of variance (7, pp. 267–273) for Trials 26–48, inclusive, when all groups were working against Load 0 (from the last point in Fig. 6 to the end of the 4-min. period), yielded a significant F for neither the Groups mean square ($df = 6$ and 633) nor the Groups × Trials interaction mean square ($df = 132$ and 13,926). The null hypothesis with respect to the effect of duration of preshift practice against Load 4 is, thus, tenable for the test performed.

Discussion

Rate recovery after rest in Exp. I and II was shown to be independent of work loading, even though absolute rate of responding was contingent upon the loading prevailing at any moment of practice. For the loading-shift Exp. III and IV, however, the data are more difficult to summarize with one general rule because statistical significance was not usually found, though the ranking of load means was typically unidirectional. Our best guess is that the loading used in present practice will account for more of cranking behavior than will the loading used in previous practice. Under certain specifiable circumstances the loading used in previous practice appears to affect present rate of responding to some small, but theoretically significant degree, while for certain other circumstances of practice, the loading used in previous practice appears irrelevant to present rate of responding. But, as will become apparent, the task of accounting for the set of fairly consistent findings is much more difficult than is the task of describing functional relationships.

Two extreme interpretations may be made of the differences between the rates of responding of groups working against different work loadings: (*a*) that the differences are entirely attributable to differences in I_R, and (*b*) that the differences are not even in part attributable to differences in I_R, but, instead, reflect an adjustment in rate of responding which exactly counteracts the differences in the I_R increments per response. The data of the present studies thus bear on the question, raised in a number of previous studies (1, 2, 4), of how the amount of I_R accrued in a self-paced task varies with the work loading of the unit response. This question has recently been given an extended discussion and review by Ellis (5).

Recovery, pre- and postrest loading constant.—The present studies of recovery in rate of responding, as well as a previous study with the same task (2), indicate the recovery is independent of work loading. A study using a different task, moreover, yielded little or no trend in magnitude of recovery as work loading increased (1). The constancy of the recovery is thus fairly well established across tasks and for a fair range of durations of practice and rest. Experiment II, in addition, yielded no between-loading differences in recovery over four successive intertrial rests. The gain measure used in Exp. I and II, of course, reflects both the recovery over rest and the decrement incurred in the first scoring period of postrest practice. If the constancy of the recovery is an artifact in which unequal dissipations of I_R during rest are counterbalanced by unequal accruals of I_R at the start of postrest practice, the artifact is remarkably consistent over a variety of experimental manipulations.

If the accrual and dissipation of equal amounts of I_R by different loading groups are defined, respectively, in terms of equal decrement and increment in number of responses per unit time, regardless of work loading, then the increase in between-loading differences in early trials indicates that, initially, I_R accrues more rapidly the greater the work loading. That the differences are fairly stable after the first minute or two of prerest practice indicates that the rate at which I_R is accrued is about equal for different loadings, though there are between-

loading differences in amount of I_R present at any time. The constant gain in response rate with rest, however, indicates, inconsistently, that equal amounts of I_R are dissipated by all work-loading groups. If the rate of dissipation of I_R is assumed to increase as the amount of I_R present at the start of rest increases, the constant gains over the rest contradict expectation based on the existence of prerest differences in rate of responding.

An alternative, that equal amounts of I_R are accrued by all loading groups, and that the equality in I_R should be evident in inequality from the start of practice in rate of responding and in rate of decrement, suffers similar inconsistencies. If it is assumed that, at the start of practice, habit and drive are constant for all loadings, it is possible that the equal ${}_S\hat{E}_R$'s might act to determine equal energy expenditures between loadings, rather than responses made with equal temporal characteristics. The measure, horsepower required to turn the crank, is not an adequate expenditure measure, however, since the power required to turn the hand and arm is not included. Presumably, however, equal gains in ${}_S\hat{E}_R$ would result in smaller gains in rate of responding, the greater the loading, according to an equal-expenditure position; thus the equal recovery in numbers of revolutions suggests unequal gains in ${}_S\hat{E}_R$ (and unequal dissipation of I_R, provided ${}_SH_R$ and D were constant at the start of practice). Greater rates of decrement in number of revolutions in prerest practice for heavier loadings (or equal rates of decrement) are also apparently inconsistent with the position that expenditure is constant between groups, since even an equal number of revolutions lost, per constant interval of time, indicates an unequal change in ${}_S\hat{E}_R$.

From the authors' point of view, at least, the recovery studies pose ambiguities in interpretation in which accepting either of the originally proposed extreme interpretations of work-loading differences leads to its rejection. For the present, the loading shift technique would seem to offer a more clear-cut answer to the question of whether differences in I_R develop between different loading groups in a self-paced motor task.

In the recovery studies, nonetheless, are two noteworthy relationships for the present self-paced task, as opposed to paced tasks. These are the relationships of rate of responding and of horsepower recovery to work loading along the load abscissa defined in the preceding section. The number of revolutions per trial appears a linear function of load (data not presented), the slope varying with the ordinal number of the trial, while horsepower gain with rest seems a negatively accelerated function of load. If the present data are interpreted as indicative of between-load differences in inhibition, the positively accelerated effect of work loading (expected in a paced task) is modified when loading is varied in a self-paced task. Some change in the functional relationship of work loading and the amount of inhibition accrued in a given duration of practice would, of course, be expected in a self-paced task where the rate of responding is less for a heavy load than for a light, even though the inhibitory increment, per response, is greater for the former.

Postshift performance.—The increase in rate of responding over instantaneous shifts to the lighter loading has some interesting implications for inhibition theory. In order to account, in terms of I_R, for some such increment in rate of responding, it is necessary to assume that I_R dissipates continuously, during practice as well as during rest. If the preshift differences reflect only differences in I_R and if I_R dissipates only during rests, on the other hand, postshift performance should exhibit only a slower rate of decrement (as with shifts after fairly brief durations of practice against the heavier loading), rather than the observed, absolute increment in rate of responding. For the moment, no concern will be given to whether or not the usual jump in response rate with a shift from Load 4 to Load 0 reflects the difference between the ${}_SI_R$'s evoked by the two loadings; if ${}_SI_R$ is thus employed,

more of the preshift difference between loadings must be attributed to ${}_{S}I_{R}$ than to I_R.

In the present studies, comparisons of the postshift performance of groups changing from Load 4 to Load 0 with groups practicing against Load 0 throughout were made in order to determine whether differences in postshift performance could be attributed to any difference in the inhibitory effects of preshift loading. As in previous studies (2, 4) the weight of the data (grand mean comparisons in postshift practice) showed a slightly greater rate of responding in unshifted, control groups than in groups shifted from a heavy to a lighter loading.

In Exp. IV, where all experimental groups were shifted instantaneously from Load 4 to Load 0, every experimental group was slightly inferior to the control group in postshift practice. There was some indication, too, that the decremental effect of preshift cranking against Load 4 increased with increasing durations of preshift practice.[4] These data are consistent with the notion that some difference in inhibition attributable to preshift loading exists in postshift practice, and that the difference in inhibition is an increasing function of the duration of preshift practice. As has been pointed out previously (2), however, it is important to note that postshift differences are slight and that performance depends more upon total duration of work and present work loading than upon previous work loading.

In Exp. III, a constant duration of preshift practice, longer than any used in Exp. IV, was used. Experimental groups were changed from Load 4 to Load 0 after varying durations of rest, while control groups worked against Load 0 before and after the rest. If the control and experimental groups differed in the amount of I_R accrued by the end of the 5-min. period of preshift cranking, the between-load differences in I_R (and, presumably, the between-load difference in performance) at the start of the postshift period would decrease as the interpolated rest increased. No such trend was evident; further, the occurrence of larger postshift differences between loads after rests of 1 to 6 min. than after 0-min. rests complicates the explanation, for the finding is most out of line with the expectation of decreasing between-load differences with increasing rest.

The data of Exp. III are, however, consistent with the previous finding that the between-load difference in postshift performance (5 min. of preshift cranking) was greater after a 4-min. rest than after a 40-sec. rest (4). In the earlier study, a pacing habit was proposed to account for the increased effect of preshift loading with a longer rest, i.e., with brief rests after a long work period, considerable I_R is present, and performance decrement is primarily dependent on I_R, while after longer rests I_R is relatively unimportant and the influence of a pacing habit, acquired in preshift practice, according to the rate of responding, can be observed in performance. The present data suggest that if a pacing-response account is used, it is not applicable for 60-min. rests.

While the present data do not in every case indicate that postshift rate of responding against Load 0 depends on the preshift loading, they do lend considerable support to the position that performance does vary, though slightly, with preshift loading. The source of such differences as were observed, however, is not established; they may depend, in whole or in part, on inhibitory factors (I_R and ${}_{S}I_{R}$), or they may reflect some learning effect. That is, *S*s working against Load 0 may strengthen some habits specific to Load 0, habits not strengthened with practice against Load 4. If such is the case, however, cranking against Load 4 should presumably strengthen habits not strengthened in cranking against Load 0. Future research, involving shifts from a light to a

[4] The postshift difference is actually correlated with the preshift difference; the empirical maximum difference between Loads 0 and 4 occurred at about 90 sec. of practice for Group 120, and declined slightly thereafter. Group 60, thus, differed more from Group 0 than did Group 120 in the last trial before the shift as well as after shifting from Load 4 to Load 0.

heavy loading, may thus help to remove some of the ambiguities of the present findings. Another technique being explored at present is to allow *Ss* to determine the durations of their rests, and to compare the rests taken after a constant period of cranking for different loadings.

Summary

Four studies used a simple cranking device with basic trainee airmen *Ss* to investigate the question of the extent to which the inhibition accrued in a self-paced task varies with work loading.

Experiments I and II, holding pre- and postrest loading constant, provided evidence that recovery in rate of responding is independent of work loading across manipulation of duration of prerest work, duration of interpolated rest, and ordinal number of the recovery. The data provided no unambiguous answer to the question of how much, if at all, work loading affects the amount of I_R accrued in a self-paced task.

In Exp. III and IV the inhibitory effects of two loadings were studied by comparing *Ss* shifted from the heavier to the lighter loading with control *Ss* who worked against the lighter loading throughout. In Exp. III the variable was the duration of interpolated rest given after a constant period of preshift cranking, and in Exp. IV the variable was the duration of cranking preceding an instantaneous shift. In general, though not in every case, the postshift performance of control groups was superior to the performance of groups shifted from the heavy to the light loading. The data thus suggest that a greater amount of I_R was accrued in work against the heavier loading, though alternative explanations may be offered to account for the postshift differences.

Postshift differences, however, were slight, performance depending more upon present than past work loading; the different rates of responding for heavy and light loadings seemed, thus, to some extent to compensate for the difference in the work required per unit response.

References

1. Bilodeau, E. A. Performance decrement in a simple motor task before and after a single rest. *J. exp. Psychol.*, 1952, **43**, 381–390.
2. Bilodeau, E. A. Decrements and recovery from decrements in a simple work task with variation in force requirements at different stages of practice. *J. exp. Psychol.*, 1952, **44**, 96–100.
3. Bilodeau, E. A. Massing and spacing phenomena as functions of prolonged and extended practice. *J. exp. Psychol.*, 1952, **44**, 108–113.
4. Bilodeau, E. A. The effects of force variations before rest on rate of responding after rest. *USAF, Hum. Resour. Res. Cent., Res. Bull.*, 1953, No. 53-26.
5. Ellis, D. S. Inhibition theory and the effort variable. *Psychol. Rev.*, 1953, **60**, 383–392.
6. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
7. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

(Received March 25, 1954)

Journal of Experimental Psychology
Vol. 48, No. 6, 1954

TILT ADAPTATION AND FIGURAL AFTER-EFFECTS[1]

ERIC G. HEINEMANN AND THOMAS MARILL

Harvard University *Massachusetts Institute of Technology*

In 1933 Gibson (1) reported that a curved line tends to look less curved after long inspection and that a straight line subsequently shown in the same location looks curved in the opposite direction. Soon afterwards he reported exactly analogous effects for bent and tilted lines (2, 3). He referred to the gradual change a curved, bent, or tilted line undergoes during inspection as "adaptation" and to the effect upon a subsequently presented straight line as "negative after-effect."

Gibson held that the crucial condition for the occurrence of these effects is the deviation of the lines initially presented from norms such as "verticality" and "straightness." He argued that long inspection of a line which deviates markedly from such a norm will cause a gradual readjustment of that norm. The direction of the readjustment is such as to decrease the apparent deviation of the inspected line from its norm. Both adaptation and negative after-effect then result from the fact that the curvature, or tilt, or bentness of the lines is being judged with reference to this newly established norm.

A different theory was proposed later by Köhler and Wallach (5). According to these authors Gibson's effects are the result of "satiation," a condition which is produced by long inspection of *any* figure. This satiation causes figures to be displaced away from the boundaries of other figures, or from regions previously occupied by the boundaries of other figures. Köhler and Wallach were able to show that deviation from a norm is not an essential condition for the occurrence of negative after-effects (renamed "figural after-effects" by them). They also applied their theory to some of Gibson's adaptation effects but pointed out that much of Gibson's evidence for adaptation is ambiguous because what he called measurements of adaptation were actually measurements of negative after-effects.

New evidence in support of Gibson's interpretation of his results was advanced by Prentice and Beardslee in 1949 (6). These investigators claimed to have found adaptation of tilted lines with conditions under which a satiation effect could not have occurred. Their findings raise the possibility that many phenomena that pass as figural after-effects are the result of two independent processes, Gibson's adaptation and Köhler and Wallach's satiation.

However, Prentice and Beardslee's results can be interpreted in terms of the satiation theory. In Prentice and Beardslee's experiments *O* was "exposed to a tilted line in one part of the visual field for a period long enough to give rise to normalization[2] (if any) and then allowed to compare the appearance of that line to that of another lying in some unaffected portion of the field" (6, p. 356). The lines were drawn in India ink upon a large rectangular piece of cardboard. It follows from the theory of Köhler and Wallach that a line that is enclosed in

[1] This research was done at the psychological laboratory, Cornell University.

[2] Prentice and Beardslee's term for "adaptation."

a rectangular framework will tend to align itself with those borders of the frame most nearly parallel to it. Since the edges of Prentice and Beardslee's cardboard were vertical and horizontal, respectively, such an alignment effect, if it occurred, would be indistinguishable from an adaptation effect. Prentice and Beardslee recognized this fact but assumed that in their experiments the distance between the lines and the nearest edges of the cardboard was too large to permit the occurrence of such an alignment effect.

The principal evidence in support of this assumption comes from a control experiment in which Prentice and Beardslee cut two of the edges of the cardboard in such a fashion that they were parallel to the tilted line. Adaptation still occurred with this modification of the setup.

A more detailed consideration of the application of satiation theory to these situations shows that the control experiment is inconclusive. The two stimulus situations (main experiment and control experiment) are shown diagrammatically in Fig. 1. Consider first the situation shown in Fig. 1a. On Köhler and Wallach's assumptions, satiation is most intense where the separation between the line and the nearest border of the frame is smallest, in the regions marked B, and least intense in the regions marked A. Such a gradient of satiation should produce the apparent rotation of the line found by Prentice and Beardslee. But a gradient of this sort will also be established in the control situation (Fig. 1b) since, according to Köhler and Wallach's theory, satiation is more intense within the acute angles of the frame than within the obtuse ones.

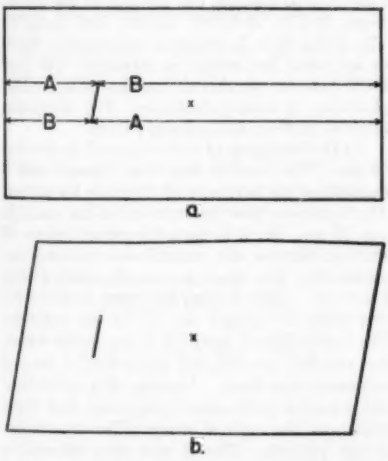


FIG. 1. Stimuli used by Prentice and Beardslee. The figure marked *a* was used in the main experiments, the one marked *b* in the control experiment. The comparison line which appeared to the right of the fixation point has been omitted.

Satiation should produce a rotation of the line relative to the edges of the cardboard as well as relative to the vertical. In the control experiment, then, the line should no longer appear parallel to the edge of the cardboard at the end of the inspection period. This is in fact what Prentice and Beardslee's *Os* reported.

The experiments reported below represent a further attempt to establish whether Prentice and Beardslee's results represent an alignment effect or an adaptation effect in Gibson's sense. Prentice and Beardslee's setup, as shown in Fig. 1a, was duplicated as nearly as possible and their experiments were repeated under two principal conditions. In Exp. I and III the inspection line was vertical and the rectangular cardboard upon which it was presented was tilted 10°. If an alignment tendency exists, the line should turn *away* from the vertical under this condition. Adaptation, of course, can have no effect in this situation since the line is already vertical, or normal.

In Exp. II both line and rectangular cardboard were tilted 10° in the same direction. Thus the line was tilted 10° with respect to the vertical but was parallel to the edges of the cardboard. An alignment effect is excluded here but adaptation, if it exists, should turn the line toward the vertical.

Method

Subjects

Nine men and three women, all graduate students in psychology, served as *Ss*. Three of the *Ss* served in two experiments each, all others in only one experiment.

Apparatus

A piece of white cardboard 22 × 28 in. was mounted on a metal stand so that it could be tilted through 10° in a clockwise or counterclockwise direction from the position in which the short edges of the cardboard were vertical and the long edges horizontal. Centered on this cardboard was a fixation mark, and symmetrically placed on each side of the fixation mark, halfway between the mark and the short edges of the cardboard, were two short black "lines." These lines were made of black metal bars 3 × ¼ × ¼ in. Each of these metal bars was mounted on a pin which was attached to the center of the bar, perpendicular to its long axis. These pins were pushed through small holes in the cardboard. A small coil spring was then placed around each pin and held in place by a clamp at the end of the pin. The springs were given enough tension to press the bars firmly against the cardboard, but not enough to prevent *E* from rotating the bars about their midpoints at will. A pointer was also attached to each pin so that the angular position of the bar could be read off on a protractor attached to the back of the cardboard.

In Exp. I and II a removable cardboard flap was used to cover one of the bars from view when this was desired. The flap was cut and mounted in such a fashion that its edges were always parallel to the edges of the cardboard.

In Exp. III the flap was not used. Instead, when only one of the lines was to be exposed the entire cardboard was covered by a second cardboard of the same dimensions. Upon this second cardboard were drawn in India ink the fixation mark and the single line to be exposed.

The apparatus was placed in front of a homogeneously-textured gray wall. The angular distance between the edges of the cardboard and the nearest vertical contour visible on the wall was approximately 30°. The average illumination for the cardboard and its surround was 5.8 ft.-candles. The illumination was nearly homogeneous and sufficiently diffuse to prevent visible shadows from the two black bars.

Observation was from a distance of 3 m. and *S*'s head was kept fixed by means of a head and chin rest. The angular dimensions were: width of cardboard—13° 32′, height of cardboard—10° 40′, length of lines—1° 27′, width of lines—3.5′.

Procedure

The basic method for measuring the alignment or adaptation effect was the same in all three experiments. The *S* was required to fixate the mark for a period of time during which only one of the lines (the standard) was exposed. At the end of this time the second line (the variable) was exposed briefly and *S* was required to judge whether its tilt was clockwise or counterclockwise relative to the standard. The procedure was repeated for different objective tilts of the variable until enough judgments were collected to compute a point of subjective equality (PSE).

In each experiment measurements were made under four conditions. The cardboard was tilted 10° to either side of vertical, and either line served as standard. Each *S* served in four experimental sessions, one for each of the conditions, R-CW, R-CCW, L-CW, and L-CCW. The initial R or L indicates whether the right- or left-hand line served as standard; CW and CCW indicate whether the cardboard was tilted clockwise or counterclockwise. The minimum interval between sessions was 7 days.

At the beginning of each session *S* fixated for 90 sec. The variable was then exposed and *S* maintained his fixation until he made his report. The maximum time for exposure of the variable was 10 sec., but if *S* made his report before 10 sec. had elapsed, the variable was covered immediately. The report was usually made within 1 to 4 sec. After making his report *S* closed his eyes while *E* changed the tilt of the variable. The *S* then fixated again for 10 sec., after which the variable was exposed again and a second judgment was made. Usually, this procedure was repeated until seven judgments had been made, one for each of seven different settings of the variable. The *S* was then allowed a recess of a few minutes until he felt rested and willing to resume the task. Four blocks of seven judgments were made in one session. The seven settings of the variable covered a range of 7° with 1° steps; they were presented in random order.

Like Prentice and Beardslee, we found that this range of stimuli was usually sufficient to give a range of judgments running from 0% to 100% clockwise. However, it was necessary twice to add more steps to complete the range of judgments. The PSE was determined by Spearman's method (4, p. 169–171).

Differences in procedure for the three experiments were as follows:

Experiment I.—The standard line was always objectively vertical and thus tilted 10° with respect to the shorter edges of the cardboard. During the inspection periods the variable was covered by the cardboard flap which was removed by *E* when exposure of the variable was desired.

Experiment II.—The standard line was always tilted 10° from the vertical in the same direction as the cardboard and thus was at all times parallel to the shorter edges of the cardboard. The cardboard flap was used as in Exp. I.

Experiment III.—This is a repetition of Exp. I, except that a full-size cardboard cover was substituted for the flap. The purpose of Exp. III was to investigate the possibility that the effect found in Exp. I could be attributed to the influence of prolonged inspection of the contours of the flap on the apparent tilt of the variable.

Results and Discussion

The results of the three experiments are shown in Tables 1, 2, and 3.

TABLE 1

Apparent Rotation of the Vertical Line in the Direction of Alignment with the Cardboard—Exp. I: Cardboard Tilted 10°

S	Line Used as Standard (R or L) and Direction of Tilt of Cardboard (CW or CCW)				Mean*
	R-CW	R-CCW	L-CW	L-CCW	
FJ	+3.25	+3.00	+2.25	+4.00	+3.13
JW	+2.75	+1.75	+2.00	+1.75	+2.06
EH	+2.25	+0.25	+1.50	+2.00	+1.50
EM	+1.00	+0.75	+0.50	+1.00	+0.81
HJ	+3.25	+2.00	+1.25	+2.50	+2.25
Mean					+1.95

* Positive values indicate an apparent rotation in the direction of alignment with the edges of the cardboard.

TABLE 2

Apparent Rotation of the Tilted Line in the Direction of Alignment with the Vertical—Exp. II: Line Tilted 10°, Cardboard Tilted 10°

S	Line Used as Standard (R or L) and Direction of Tilt of Cardboard (CW or CCW)				Mean*
	R-CW	R-CCW	L-CW	L-CCW	
JP	+0.50	+1.00	−0.25	+0.50	+0.44
HJ	−1.00	+0.25	+0.25	+0.25	−0.06
WC	+0.50	0.00	+0.75	0.00	+0.31
JC	−0.50	+1.25	+0.25	+1.00	+0.50
SH	−0.50	−0.25	−0.25	+0.50	−0.13
Mean					+0.21

* Positive values indicate an apparent rotation toward the vertical.

TABLE 3

Apparent Rotation of the Vertical Line in the Direction of Alignment with the Cardboard—Exp. III: Cardboard Tilted 10°

S	Line Used as Standard (R or L) and Direction of Tilt of Cardboard (CW or CCW)				Mean*
	R-CW	R-CCW	L-CW	L-CCW	
DW	+2.50	+1.92	+1.25	+2.00	+1.92
EH	+3.00	−2.25	+1.00	+1.00	+0.69
JB	+2.00	+2.13	+1.25	+2.25	+1.91
EM	+1.00	+1.00	+1.50	+1.25	+1.19
JR	+1.50	+0.50	+1.00	+1.50	+1.13
Mean					+1.36

* Positive values indicate an apparent rotation in the direction of alignment with the edges of the cardboard.

Each entry in these tables represents the difference, in degrees, between the angular positions of the standard and variable at the PSE. For Tables 1 and 3, differences were computed in such a way that those indicating a rotation of the standard in the direction of alignment with the cardboard appear as positive. In Table 2 those differences that indicate a rotation of the standard toward the vertical

appear as positive. The notation for "conditions" is discussed in the section on Procedure.

The one-tailed t test was used to evaluate whether the grand means are significantly different from zero. For those obtained in Exp. I and III the null hypothesis can be rejected at better than the 1% level of confidence. For the grand mean obtained in Exp. II the null hypothesis cannot be rejected even at the 5% level.

Thus vertical lines showed a significant tendency to turn away from the vertical, in the direction of becoming more nearly parallel to the edges of the cardboard. On the other hand, lines tilted 10° relative to the vertical did not, under our conditions, show any significant tendency to turn toward the vertical.

The mean effect found by Prentice and Beardslee was 2.12°. In view of the large individual differences found in experiments of this sort, the very small disagreement between their results and those of our Exp. I and III need raise no questions.

We conclude that the results of Prentice and Beardslee can be accounted for by the alignment tendency demonstrated here and do not provide evidence for an adaptation effect in Gibson's sense.

Summary

With an arrangement similar to that of Prentice and Beardslee (6) three experiments were made on the change in apparent tilt of lines after long inspection. In Exp. I and III the standard line was always vertical and the cardboard upon which the standard was presented was tilted 10° either clockwise or counterclockwise. In Exp. II the standard line and the cardboard upon which it was presented were both tilted 10° from the vertical in the same direction. Thus, although in Exp. II the line was tilted 10° relative to the vertical, it was always parallel to two of the edges of the cardboard.

In Exp. I and III the objectively vertical lines showed a significant tendency toward alignment with the edges of the cardboard, i.e., a tendency to change their apparent orientation so as to become more nearly parallel to two of the edges of the cardboard. In Exp. II it was found that there is no significant tendency for a line to appear to turn toward the vertical when an alignment effect is excluded.

It was concluded that the results of Prentice and Beardslee represent an alignment effect and not adaptation to the norm in Gibson's sense.

References

1. Gibson, J. J. Adaptation, after-effect and contrast in the perception of curved lines. *J. exp. Psychol.*, 1933, **16**, 1–31.
2. Gibson, J. J. Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. *J. exp. Psychol.*, 1937, **20**, 453–467.
3. Gibson, J. J. Adaptation, after-effect, and contrast in the perception of tilted lines. II. Simultaneous contrast and the areal restriction of the after-effect. *J. exp. Psychol.*, 1937, **20**, 553–569.
4. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
5. Köhler, W., & Wallach, H. Figural after-effects: an investigation of visual processes. *Proc. Amer. phil. Soc.*, 1944, **88**, 269–357.
6. Prentice, W. C. H., & Beardslee, D. C. Visual "normalization" near the vertical and horizontal. *J. exp. Psychol.*, 1950, **40**, 355–364.

(Received March 27, 1954)

Journal of Experimental Psychology
Vol. 48, No. 6, 1954

THE EFFECT OF TRAINING ON ABSOLUTE ESTIMATION OF DISTANCE OVER THE GROUND[1,2]

ELEANOR J. GIBSON AND RICHARD BERGMAN

Cornell University

Does the judgment of distance depend on learning? This question has been of theoretical interest to psychologists for many years, but its practical importance is equally evident. Even with advances in instrumental control of ranging and aiming, the necessity of human judgments of distance in many military operations has not been eliminated. Previous studies have shown that training can improve absolute estimations of the distance of a target from the observer (**4, 5, 7, 8, 9**). The conditions under which greatest improvement will occur and the relative specificity of the resulting skill are questions in need of further investigation.

The training methods and situations used, to date, may have resulted in rather specific cue-response associations. In two studies (**4, 8**) *Ss* were given training in making absolute estimations, in yards, of the distance to aerial targets. Since distance cues for an object seen through the air are nonexistent, except for those in the object itself, judgment was probably based primarily on the apparent size of a familiar target (airplane). The *Ss* probably associated a certain perceived size of the airplane with the number of yards called out by *E* during training trials. Such training might have limited transfer value to unfamiliar targets or new field conditions.

On the other hand, when the target is located on the ground, the surface stretching between *S* and target would provide perspective and texture gradients in correspondence with varying distance stretches.[3] Training under these conditions might yield more generalizable improvement, since cues independent of any particular target would be available. One study with targets on the ground (**5**) has been performed, but since the targets were familiar existing landmarks, such as telephone poles, it is conceivable that improvement again was linked to the specific targets.

That *Ss* are likely to hit upon rather specific cues during practice and achieve improvement by associating them with those responses which are reinforced was suggested by an experiment by Gibson and Smith (**1**). It was proposed to measure transfer of training of distance estimations to size constancy judgments (which would presumably be an alternative measure of distance judgment). The experimental group was given training in estimating distance (in yards) to pictured targets (stakes) in photographs. Size judgments were then obtained for the stakes, following the matching procedure used in a previous study (**2**). A control group made the size judgments without any pretraining. No significant transfer was found by measures of constant or average error, though group variability was reduced by the training. The *Ss* showed a strong tendency to form highly specific associations between particular identifying clues in the photographs and a verbal response of so many yards. This accomplishment, which the correction procedure of the training encouraged, was apparently irrelevant for the judgment of the size of an object at a given distance.

These considerations suggested that the crucial question to be explored is whether it is possible to train *Ss* in judging distances in such a way

[1] This research was supported in part by the United States Air Force under Contract No. 33(038)-22373 monitored by the Perceptual and Motor Skills Research Laboratory, Human Resources Research Center, Lackland Air Force Base. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government. A more detailed report will be available in a Research Bulletin of the AF Personnel and Training Research Center series.

[2] Mrs. Noel Naisbitt, Mr. Alfred Steinschneider, and Mrs. Naomi Wolin assisted with the statistical analysis at various stages of the project. Miss Jean Graham drew the charts. Thanks are due Dr. T. A. Ryan for advice on the choice of statistical procedures.

[3] See Gibson (**3**) for a discussion of the role of surface in space perception.

that generalizable improvement is obtained. If there are changes from training to test in the target objects, stimulus values, or type of judgment required, can any transfer of improvement be expected? The experiment to be described was planned to determine whether training can improve the absolute estimation of distance over the ground when a different stretch of distance from *S* to the target is to be judged on every trial. If no opportunity is given to memorize the distance in yards to any target, will training by corrective reinforcement still result in improvement of estimation of other distance stretches? In other words, the distance stretches presented as stimuli for judgment during the training series were not repeated in the test series, so that it could be determined whether any skill had been achieved which would transfer to the judgment of different distance stretches.

The intention was to limit the potential cues to those provided by the ground surface, insofar as possible, rather than to permit the target itself to provide the cues for judgment. Consequently, the target could not be unique for each judgment, and the kind of terrain chosen for the judging ground was most important. We used a level athletic field, 132 × 488 yd., with a surface of mown grass. It was free of all objects except the targets, which were metal rectangles mounted on stakes. The field was bounded at the lower end, where *S*s were stationed, by a terrace and a group of buildings. They looked toward the upper end, bounded by trees and a gymnasium. The 21 targets were arrayed permanently on the field, but none stood in line with another.

As *S* looked toward a target to make his judgment, he viewed a stretch of grass of a given distance bounded by a target with no particular contextual or associative value. There was no interposition, no shadows, and no linear perspective. Size perspective, in the sense of gradients of texture density on the ground surface, was present, and might have been enhanced by binocular disparity and motion parallax. Some of the other targets were visible in the periphery and might also have provided gradients of size perspective. The targets varied in up and down location in the field, since a horizon was clearly visible. The question is, could *S* improve his skill in making absolute estimates by responding to variables of texture density and size provided by the gound and target array, plus vertical location of the target, without memorizing specific cue-yard associations?

Method

Arrangement of targets.—The principal problem in planning the experiment was to arrange the targets on the available ground so as to provide different distance stretches for every training trial, and for the test series. Six observer station points were located in the lower right hand corner of the field (indicated on the map, Fig. 1, as 1, 2, 3, 4, 5, and 6). The 21 targets (A to U on Fig. 1) were arrayed over the field in such a way that a large number of different distance stretches would be provided when they were combined variously with the six station points. One-hundred and eight stretches were chosen from the total possible number so that no target ever stood in a line between *S* and the one designated for judgment, and so that there was a fairly even distribution over the total range. True distances to the targets which *S* was asked to judge averaged approximately the same at all station points. Station points and target loci were staked out by a surveyor and were accurate to a few inches.

Targets.—It would have been desirable to use a single target, which could be moved to the various target loci. This was impossible, due to the large size of the field, the necessity of accurate location, and the fact that two *S*s were sometimes run at once. The targets had to be permanent and stationary, and they had to differ so that they could be indicated to *S*.

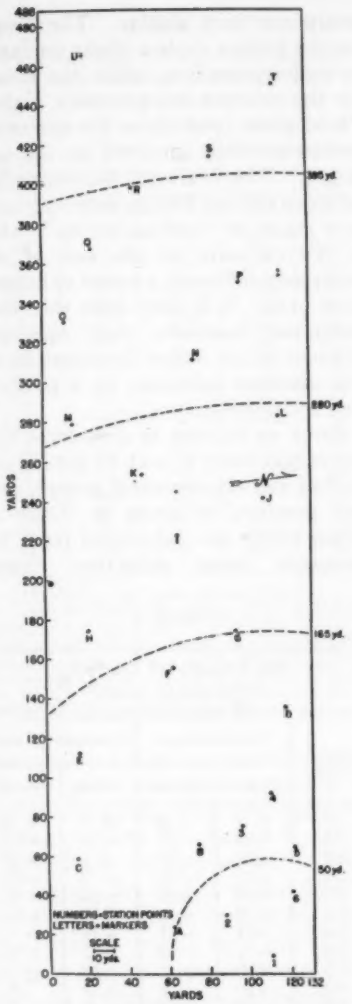


FIG. 1. Map of field, showing station points (1 through 6) and target locations (A through U)

Labels by number or letter were not possible, since even very large ones could not be read 400 yards away. Furthermore, some colors were difficult to identify at a distance. A 3-plate, 3-color system of target marking was therefore chosen. One to three metal plates, 18 × 24 in., were attached to a wooden stake. These plates were painted white, yellow, or black. A plate was always fastened at the bottom of the stake, close to the ground, since a raised target tends to "ground" itself and may, by appearing higher in the field, be increased in apparent distance. All the targets were readily identifiable from the station points.

Plan of experiment.—Two groups of *S*s, an experimental group and a control group, were given a pretest at Station Point 1 (18 judgments of distance, in yards, to targets ranging from 52 to 395 yd.) and a posttest with the same distance stretches. All these judgments were uncorrected. The experimental group received training between the two tests, consisting of 90 differing judgments of the distance of targets ranging from 39 to 435 yd. (targets viewed from Station Points 2, 3, 4, 5, and 6). After *S* had made his estimate, it was corrected by *E*. The control group received no training, but spent the interval between tests taking a paper and pencil test of mechanical aptitude.

Since the pretest and posttest utilized the same 18 depth stretches and targets, two random orders were made. Half of the *S*s judged in order A for the pretest and B for the posttest; and half the reverse arrangement. The control group and experimental group were divided in the same way. The experimental group was also divided into two subgroups which differed in the order of station points during the training series. This alternative sequence of station points was not thought necessary as a control, but was adopted because it permitted two *S*s to be run at the same time without one *S* and his *E* ever obstructing the view of the other *S*. The group of 18 judgments made at a station during the training will be referred to as a training trial. Since five station points were employed during training, there were five training trials.

Procedure.—The *S* was first asked to read a Snellen chart, and then given his instructions. These included practice in identifying pictured targets using the same 3-plate, 3-color system as those on the field. He was told that he was to estimate how far away from him a target was, in yards, but no demonstration of a yard was given. He was told that he would judge the same target more than once, but never from the same station point. The *S*s in Group E, who were given correction, were further told that there was no point in memorizing any one figure because no distance stretch would be exactly repeated.

Records were kept of all of *S*'s estimates, in yards, as well as his age, civilian occupation, and potentially relevant experience (such as track and football). The time of day and the weather were also recorded. Two check lists for weather were utilized, a descriptive one (sunny, sunny hazy, cloudy scattered, cloudy

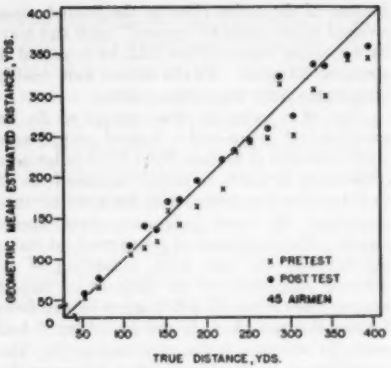


FIG. 2. Function relating estimated distance to true distance (data from Group E, airmen)

broken, cloudy overcast) and a visibility rating. Subjects were not run when it rained.

Subjects.—A group of 30 *Ss* drawn from the Cornell population was run first. These people varied in age, sex, and academic background. A second group of 92 *Ss* was drawn from the group of airmen in basic training at Sampson Air Force Base. They were highly motivated and obeyed instructions to the letter.

RESULTS

Analysis by Target Distance

The *Ss*' estimates of distance to the 18 targets on the pretest and posttest were transformed to four-place logarithms, since previous studies (**4, 5, 6**) have shown the desirability of such a conversion. Frequency distributions of estimates in yards were skewed, since the yard scale used by *S* was "open-ended" at the top. The logarithmic transformation yielded reasonably normal distributions and also permitted use of geometric means of the estimates.

When the estimates are plotted as a function of true range, the relationship is seen to be linear. The geometric mean estimates plotted in Fig. 2 are taken from the pretest and posttest of Group E, airmen. The plots for Group C, airmen, and for the Cornell group are very similar. The means for the pretest show a slight tendency to underestimation, while the means for the posttest are generally higher. These plots (and those for the other groups as well) level off at the last target. The *Ss* are not discriminating between the last two targets—perhaps as a result of "end-anchoring"—that is, *S*'s estimate of the end of the series may influence a target or targets close to it. It is clear from the other estimates, however, that apparent distance of an object increases as its real distance increases, in a predictable relationship.

Error in relation to distance.—The percentage error at each target for the control and experimental groups, pre- and posttest, is given in Table 1. These errors are calculated from *Ss*' geometric mean estimates. Signs

TABLE 1

PERCENTAGE OF ERROR OF *Ss*' ESTIMATES FOR TARGETS AT VARYING DISTANCES (AIRMEN)*

Target Distance	Control Group		Experimental Group	
	Pretest	Posttest	Pretest	Posttest
52	− 7	0	+ 3	+ 9
68	−14	−17	− 2	+12
107	−15	−18	− 1	+ 8
126	−11	−11	− 9	+12
139	−19	−19	−14	− 1
154	− 9	− 9	+ 3	+11
166	−18	−11	−14	+ 4
188	−11	−10	−13	+ 4
219	−10	− 8	−14	+ 1
232	−10	− 5	+ 3	0
252	− 7	− 8	− 3	− 2
274	−15	− 5	− 9	− 5
287	+ 5	+ 3	+ 9	+12
304	−18	−12	−17	− 9
328	+ 1	+ 1	− 6	+ 2
342	−10	− 8	−12	− 2
369	−12	− 9	− 7	− 6
395	−13	−13	−13	− 9

* The error of the geometric mean of the estimates (in log scores) from the true range has been used to obtain the percentage of error; the antilogarithm of the error (the ratio of the geometric mean estimate of range to the true range) × 100 yields the percentage of error. The statistical method followed is essentially that described in reference 6.

TABLE 2

PERCENTAGE OF *SD*'s OF *Ss*' ESTIMATES FOR TARGETS AT VARYING DISTANCES (AIRMEN)*

Target Distance	Control Group		Experimental Group	
	Pretest	Posttest	Pretest	Posttest
52	69	55	58	29
68	77	58	72	29
107	77	60	112	25
126	85	62	81	27
139	74	66	97	25
154	85	60	124	20
166	75	73	85	21
188	72	65	64	20
219	76	76	67	20
232	85	71	135	15
252	86	80	139	21
274	89	72	86	19
287	88	69	135	20
304	90	76	36	22
328	85	69	99	17
342	77	80	85	17
369	87	71	111	22
395	82	76	112	19

* The *SD*, in logarithms, is converted into a ratio by finding its antilogarithm; this, times the geometric mean, gives a score one *SD* above the mean; and, divided into it, one below. For instance, the antilog of the *SD* for the control group at 52 yd. is 1.69. The geometric mean is 49; the score one *SD* above is 83 yd., whereas the score one *SD* below is 29 yd. Values in the table have been converted to percentages by subtracting 1 and multiplying by 100.

were kept, so these are constant errors. The errors are predominantly of underestimation, with the exception of the posttest for Group E, which followed the training series. All but one target in this group's posttest shows a shift in error away from minus and toward plus. No such trend appears for the control group, which had no interpolated correction. A notable feature of these results is the absence of any tendency for the direction or magnitude of the percentage error to vary with the distance. Similar calculations for the Cornell sample showed the same trends—a shift away from underestimation as a result of training, and independence of distance and magnitude of percentage error.

Variability in relation to distance.—Table 2 gives percentage of *SD*'s of estimates of the various targets. The most notable feature of these *SD*'s is their size; the variability of the *Ss*, before training, was enormous. They were, actually, very different; one airman, an 18-year-old Puerto Rican, said that he had never seen a yardstick. The effect of the training is to reduce the *SD*, as the figures for the posttest of the experimental group clearly show. Variability drops slightly for the control group, probably as a result of somewhat greater consistency of individual judgments, as will be shown later. The pretest figures and the posttest for the control group show no consistent tendency for percentage of *SD* to vary with distance. But there is some tendency for the *SD* to decrease as distance increases, after training. This tendency is confirmed by the data for Group E of the Cornell sample. Since it is most evident for the last few targets, it may again be a function of end-anchoring; i.e., the farthest point on the scale becomes a well-defined reference point and causes *Ss*' judgments near the end of the field to center closely around an estimate associated with the end of the distance scale.

Analysis by Subject

Accuracy.—Since the principal object of the experiment was to determine whether training brought about improvement, a comparison of individual error and variability before and after training is of primary interest. Of the 45 *Ss* in Group E, 41 improved; that is, the mean constant error is reduced, whatever its direction. Of the 47 *Ss* in Group C, 26 improved. The percentage of *Ss* improving in Group E is 95.6; in Group C, 55.3.

The figure for Group C is not significantly different from chance. The two percentages are significantly different from each other ($p < .0001$). If the mean net gain is determined by subtracting the amount of decrement from the amount of improvement and dividing by N, the resultant figure is nearly 16 times as great for Group E as for Group C (mean net gain = .1625, in logs, for Group E and .0128 for Group C).

It is interesting to compare the types of change, also, for the two groups. In Group E, 43 of the 45 *Ss* shift *away* from the previous constant error; for instance, an underestimator may shift to overestimation, but he is very unlikely to enhance his error of underestimation. In the control group, however, the original constant error was actually enhanced in 17 out of 47 cases. If the frequency of shift *away* from the previous CE is compared in the two groups by the chi-square test the difference is significant at the .001 level.

Variable error.—A question of major theoretical interest was whether or not training caused *Ss* to become more consistent in their judgments, whatever the tendency to over- or underestimate. Does the variability of the estimates of the individual *Ss* decrease with corrected practice; or, for that matter, with uncorrected practice? To determine variability for individuals a formula[4] which corrected for man and target mean was employed since distance stretches were not repeated. It is clear that variability was reduced in the experimental group; the mean percentage of *SD* was 34.6 before training and 18.3 after it. There was a small reduction for the control group also; the mean dropped from 29.4% on the pretest to 25.5% on the posttest. The significance of these differences was tested by Wilcoxon's (**10**) nonparametric test for paired replicates. The difference between pretest and posttest for Group E is significant at the .001 level; that for Group C at the .04 level (using a one-tailed test). The training which Group E received was therefore very effective in reducing individual variability. The practice which Group C obtained on the pretest alone, without any special training or correction, seems also to have produced a small reduction in variability of judgment. Since constant errors did not shift toward any common norm in this group, it seems likely that this reduction in individual variability, even though small, accounted for the similar small reduction in group variability.

Analysis of variance.—In order to check some of these trends, an analysis of variance was performed on the data yielded by the Cornell sample, Group E. Table 3 presents the results of these analyses for pretest and posttest. The estimates have been converted into log scores, as in the other sample. Our principal interest, again, was in a comparison of the pretest and posttest, between which training was given. The group variability (variance due to men) was reduced, as in the airman population. And the individual variability (variance due to man × target)[5] was reduced, likewise, from .0157 to .0056. This difference

[4] For each *S*, variability was determined by computing $(X - \bar{X}_M - \bar{X}_T + \bar{\bar{X}})^2$ in which X = *S*'s estimate on a given target, $\bar{X}_M$ = the mean of all *S*'s estimates, $\bar{X}_T$ = the mean for the given target, and $\bar{\bar{X}}$ the grand mean.

[5] This is the error term. It is assumed that this term reflects degree of individual variability since variance due to individual differences and to targets has been removed. Experimental error should remain constant from pretest to posttest, so that a difference between the error terms in the two tests would indicate a change in the factors producing within-*S* variance.

TABLE 3

ANALYSIS OF VARIANCE OF DISTANCE ESTIMATES (IN LOGS) MADE BY CORNELL *Ss*, GROUP E

Source	*df*	Pretest		Posttest	
		Mean Square	*F*	Mean Square	*F*
Order	1	.0306	.0207	.1186	1.1455
Between *Ss* in same group	18	1.4778		.0983	
Targets	17	.0216	1.3758	.0204	4.0000
Order × Target	17	.0260	1.6561*	.0143	2.5536**
S × Target	306	.0157		.0056	
Total	359				

* $p = .05$.
** $p < .01$.

is significant at better than the .01 level, confirming the finding that within-*S* consistency is increased by training.

The analysis of variance permits, also, an examination of the effect of order of presentation of targets. Order as such does not produce significant variance, and on the pretest, order-target interaction is barely significant at the .05 level. But after training, the variance due to this source is clearly apparent, perhaps because other sources of variance which masked it before have been reduced. It may be that *S*, in judging, always decided first whether the distance being judged at present was longer or shorter than the just previous one, and perhaps even what fraction or multiple it was. The size of the distance stretch or stretches preceding a given judgment, therefore, would have an effect on any given judgment; in fact, a different effect depending on the relative farness or nearness of the particular target currently to be judged.

Course of Learning

Having established the effectiveness of training for absolute judgments of distance, the question of the course of the learning arises. What kind of learning curve is exhibited by these changes? Since the calculation of geometric means was too laborious to be carried out for all the training trials, medians were employed for studying the trend. These medians were calculated without regard to sign, since plus and minus errors might otherwise balance each other out and give the appearance of no error. Figure 3 shows a plot of the median error expressed as a percentage of true distance for the targets of the pretest combined, for each of the training trials consecutively arranged, and for the posttest. A training "trial," in this case, means all 18 judgments made at a given station point. The first and third quartiles are included

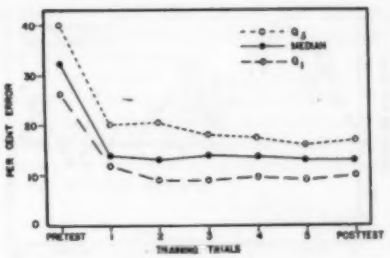


FIG. 3. Course of learning as measured by crude median error (data from Group E, airmen)

in the plot also. The median error before training is high (33%), but on the first training trial it is reduced by more than half (to 14%). No further reduction occurs, except that the quartiles tend to approach the median more closely in later training trials. The most interesting aspect of these curves is the percentage of error for the posttest. Since no further correction was given on the posttest and since none of the judgments duplicated any made during training, it might have been supposed that the error would rise again. No such tendency occurred, however. The effects of training persisted undiminished, in spite of nonidentity of the stimuli in the test series with any of the training ones. Our original question is thus answered positively: improvement in absolute judgment is possible, when distances are judged over ground, even though identical judgments are not repeated and no opportunity for memorization occurs.

Learning trends for over- and underestimators.—What happens to persistent constant errors of individuals, when their judgments are corrected? To answer this question the airmen, Group E, were divided into over- and underestimators on the basis of their geometric mean errors on the pretest (these were constant errors). For each subgroup the median error was calculated as a percentage of true distance, retaining the signs. These errors are plotted in Fig. 4 for each judgment of the pretest, each judgment of the first training sequence, and the posttest judgments. The judgments are consecutively arranged in the order in which *S* was asked for them. The line drawn through the center represents zero constant error. During the pretest the overestimators show a consistent plus error for every judgment. The underestimators, on the other hand, stay consistently minus below the zero line. But on the second judgment of the first training sequence (after *one* correction), the curve for the underestimators rises, and that for the overestimators drops. There are, of course, fluctuations, but by the ninth training judgment the fluctuations begin to parallel one another for the two groups. The curves, hereafter, even in the posttest, are remarkably similar. The training does, therefore, cause a change in constant error. This change occurs very rapidly.

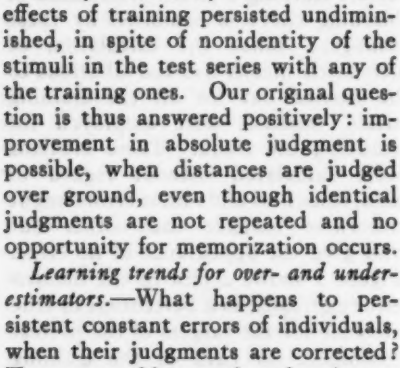


FIG. 4. Changes in constant error as a function of training

DISCUSSION

The principal question of the experiment has been answered in the affirmative: absolute judgment of distance can be improved by training, even when no opportunity is given for memorizing the distance to particular targets. Under the conditions of the experiment there is transfer of training to the judgment of "new" stretches of distance. These are new in the sense that the distance from *S* to the target is not the same as that given as a stimulus for judgment in any training trial. (The target marking the end of the stretch has appeared before, during training, but is irrelevant as a cue, since it marked different lengths of ground during training.) The ground surface itself provided a basis for consistently varying cues (texture per-

spective, retinal disparity, motion parallax, vertical placement) and these *dimensions* of variation were present for every judgment. What *S* learned was to evaluate points along the relevant stimulus dimensions in accordance with a yard scale so that he was able to *identify* a given distance stretch by naming a quantity of yards.

The hypothesis will be advanced that what is learned here can best be described as a *conceptual scale* of yards in a psychophysical relationship with stimulation provided primarily by a receding stretch of ground. This scale might be characterized by having:

1. a unit (a yard, or more likely some convenient and rounded multiple of yards);
2. boundaries or anchors at both ends (the near anchor, zero, is given by *S*'s own position, and the far anchor is acquired during the training; perhaps it is defined for *S* as the longest distance ever mentioned by the experimenter during training);
3. differentiated areas within the scale, which are fractions of the total and multiples of the basic unit (the *S* fractionates within his scale, we assume, when making an estimate in yards for a target at some intermediate point along the scale).

Does this hypothesis of a conceptual and perceptual scale explain the measurable changes brought about by learning? It certainly explains the correction of constant errors, for either a far anchor or the yard unit would provide a basis for shifting from any bias caused by an inaccurate scale unit. But what of the reduction in variable error found in comparing individual *SD*'s on the pre- and posttest? It would appear that *S* became more discriminating in his assignment of categories within the scale. This might occur because he acquired a consistent unit and end anchor during the training to use in identifying or labeling the distance stretches presented for judgment. His concept of yard might have fluctuated before the training sessions presented him with an external anchor. However, it is also conceivable that he learned not only to *identify* consistently (by naming yard quantities), but also improved in the *ordering* of points along the stimulus dimension, whatever the categories used. This implies an improvement in discrimination—a better differentiated *perceptual* scale. Whether the ordering of stimuli, as well as categorizing, was affected by the training is a problem for further investigation.

Summary

Previous experiments have shown that absolute estimations of distance, in yards, can be improved, but there is reason to believe that these training situations resulted in improvement which was specific to the targets and other conditions of the training. The present experiment was designed to see whether training would result in improvement when the targets themselves provided no cues and when memorization of specific cues and yard numbers was not possible.

Targets were viewed over a mowed grass surface from six station points. Twenty-one targets were so placed that 108 judgments, all of different distances, could be made by rotating station point and target combinations. The experimental group judged the distance, in yards, to 18 targets (pretest), then made 90 corrected judgments (training series), and finally repeated the pretest. The control group performed the two test series without intervening practice.

Results suggested the following conclusions:

1. Improvement in absolute judgment of distance occurred as a result of training, even though none of the distances presented for judgment were repeated.
2. Training tended to correct constant errors of both over- and underestimation in the experimental group. Learning curves showed that this shift occurred very quickly.
3. Variable error was reduced from pre- to posttest in both the experimental and the control group, but the reduction was greater for the experimental group.
4. The function relating true distance to estimated distance was shown to be linear.
5. The development of a conceptual scale of distance in a psychophysical relationship with stimulation provided by a receding stretch of ground was hypothesized.

References

1. Gibson, E. J., & Smith, J. A. The effect of training in distance estimation on the judgment of size-at-a-distance. *USAF*,

Hum. Resour. Res. Cent., Res. Bull., 1952, No. 52-39.

2. GIBSON, J. J. (Ed.) Motion picture testing and research. Washington: U. S. Government Printing Office, 1947. (*AAF, Aviat. Psychol. Program Res. Rep.* No. 7.)
3. GIBSON, J. J. *Perception of the visual world.* Boston: Houghton Mifflin, 1950.
4. HOROWITZ, M. W., & KAPPAUF, W. E. *Aerial target range estimation.* (OSRD, 1945; Publ. Bd., No. 15812.) Washington: U. S. Dep. Commerce, 1946.
5. PRINCETON BRANCH, FIRE CONTROL DIVISION, FRANKFORD ARSENAL. Analysis of range estimation data. Branch Memorandum No. 20, 1943.
6. PRINCETON BRANCH, FIRE CONTROL DIVISION, FRANKFORD ARSENAL. Visual range estimation, direct and differential. Branch Memorandum No. 34, 1943.
7. ROGERS, M. H., SPROL, S. J., VITELES, M. S., VOSS, H. A., & WICKENS, D. D. *Evaluation of methods of training in estimating a fixed opening range.* (OSRD, 1945; Publ. Bd., No. 4021.) Washington: U. S. Dep. Commerce, 1946.
8. VITELES, M. S., GORSUCH, J. H., BAYROFF, A. C., ROGERS, M. D., & WICKENS, D. D. *Learning range estimation on the firing line.* (OSRD, 1945; Publ. Bd., No. 4023.) Washington: U. S. Dep. Commerce, 1946.
9. VITELES, M. S., GORSUCH, J. H., & WICKENS, D. D. *History and final report of project N-105.* (OSRD, 1945; Publ. Bd., No. 4018.) Washington: U. S. Dep. Commerce, 1946.
10. WILCOXON, F. *Some rapid approximate statistical procedures.* Stamford, Conn.: American Cyanamid Co., 1949.

(Received March 26, 1954)

Journal of Experimental Psychology
Vol. 48, No. 6, 1954

S–R COMPATIBILITY: CORRESPONDENCE AMONG PAIRED ELEMENTS WITHIN STIMULUS AND RESPONSE CODES [1]

PAUL M. FITTS AND RICHARD L. DEININGER

The Ohio State University

The present paper is the second dealing with the usefulness of the concept of S-R compatibility for behavior theory. This concept concerns the effects of a class of variables that influence task difficulty in experiments in which learning, complexity (amount of information), and discriminability are controlled. Compatibility effects are conceived as resulting from hypothetical information transformation processes (encoding and/or decoding) that intervene between receptor and effector activity. The rate of processing information is assumed to be maximum when these recoding processes are at a minimum.

The objective of the study of compatibility effects is to discover conditions under which these effects occur, and to establish principles that will permit specification of the nature and difficulty of perceptual-motor tasks in terms of (hypothetical) intervening information transformation processes. Such processes must be inferred, just as do constructs such as habit strength, from measures of performance obtained in appropriate experiments. The type of experiment of greatest interest for the present purpose is one in which it is possible to measure the rate of information transfer as a function of (*a*) the choice of stimulus sets, (*b*) the choice of response sets, and (*c*) the method of combining the elements of stimulus and response sets to form S-R ensembles, and in particular to evaluate the interactions among these three variables.

The concept of intervening information transformation processes permits a reformulation of numerous problems that have stimulated a good deal of previous research. Among these are the problems of control-display relationships (**1, 5, 7, 9, 12, 13, 14, 15, 16**), the effects of stimulus-response reversals (**2, 6, 11**), and certain aspects of such classical problems as similarity and discriminability.

The results of the previous study in this series (**4**) supported the conclusion that the rate at which the perceptual-motor system can process information is a function, not so much of the characteristics of a particular set of stimuli or of a particular set of responses, but rather of the degree to which the sets of stimuli and responses form a congruent match. In the earlier study all stimuli and all responses were spatial in character, and all stimulus-response pairings were made to agree as closely as possible with population stereotypes. In the present study another S-R coding variable is investigated, the degree of congruence in the pairings of the elements within stimulus and response sets.

Specifically, the present experiment was planned to test the hypothesis that one of the conditions necessary for

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(038)-10528 with The Ohio State University Research Foundation, monitored by the Air Force Personnel and Training Research Center, and was reported at the 1953 meeting of the Midwestern Psychological Association. Permission is granted for reproduction, publication, use, and disposal in whole or in part by or for the United States Government.

maximum S-R compatibility (i.e., for maximum rate of information processing) is that the pairings of stimulus and response elements agree with a strong population stereotype. Clearly, the concept of population stereotype is related to that of S-R compatibility. However, the two concepts are defined by independent operations—population stereotypes by determination of the relative frequency with which each permissible response is made to each stimulus in a situation in which *E* gives no indication of what is considered an appropriate response (**3**, p. 1306), and compatibility effects by determination of the relative rates of gain of information when *S*s must conform to particular stimulus and response coding conditions. In the present experiment three arbitrary levels of agreement with population stereotypes—maximum, mirrored, and random S-R pairings—were employed in forming S-R ensembles.

The present study was also designed to permit a further test of the compatibility effect examined in the previous study (**4**), namely, variation in the degree of compatibility as a function of correspondence of the dimensions employed in forming stimulus sets and response sets. Two types of stimuli, spatial (lights) and symbolic (numbers and letters), were matched with a single set of (spatial) responses, and it was hypothesized in this case that greater compatibility could be achieved by the use of spatial than by the use of symbolic stimuli. The response set employed in the present study was identical with one of the response coding sets employed previously and the results from the use of symbolic vs. spatial stimuli in the present instance are considered chiefly in relation to the earlier findings. The present design, using several stimulus sets, also permits an evaluation of a third compatibility effect, the interaction of stimulus sets per se with the method of assigning specific stimuli to specific responses.

Method

Apparatus.—A schematic drawing of the apparatus is shown in Fig. 1. At the start of each trial *S* placed a stylus, which he held in his preferred hand, in contact with a ¼-in.-diameter metal button located at the intersection of eight pathways that radiated from this point like the spokes of a wheel. The angle between each pair of adjacent paths was 45°. At a signal *S* moved the stylus quickly in one of the eight directions. Reaction (decision) time was measured by a 1/100-sec. timer which started at the onset of the stimulus and stopped as soon as the stylus had been moved far enough to break contact with the metal button.

Four different sets of eight stimuli were matched with this single set of eight spatial responses. Figure 2 illustrates the four stimulus sets, and the various ways in which the elements of each set were paired with the responses to form S-R ensembles. Stimulus sets were selected to vary in their degree of physical and/or learned spatial correspondence with the response set.

The first stimulus set was composed of eight lights arranged in an octagon pattern around the periphery of an 8-in.-diameter circle. Each stimulus element corresponded spatially to the termination of one of the elements of the response set. This set of stimuli will be referred to as a

Fig. 1. Schematic drawing of the apparatus. The example shows a stimulus from the symbolic set and a response, down and to the right, which represents maximum S-R correspondence for this stimulus. (The stimulus numeral is not drawn to scale.)

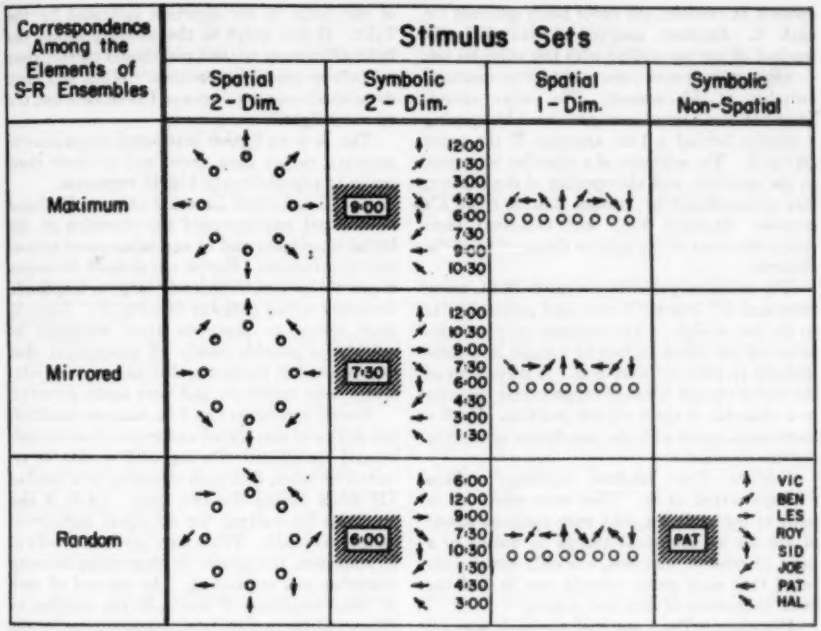


FIG. 2. The ten S-R ensembles investigated. The arrows indicate the directions of response movement designated by each stimulus. Response directions were different for each *S* in the case of random S-R pairing.

spatial two-dimensional set. It might also be called a pictorial stimulus set.

The second stimulus set, which had a learned correspondence to the response set, consisted of eight three- and four-digit numbers which were chosen to symbolize eight equally-spaced points around a clock face (1:30, 3:00, . . . 12:00). This will be referred to as a symbolic two-dimensional stimulus set.

The third stimulus set consisted of a horizontal row of eight lights spaced 1 in. apart. It will be referred to as a spatial one-dimensional stimulus set.

The fourth stimulus set consisted of eight three-letter first names. Familiar first names were used in preference to less meaningful stimuli in order to minimize stimulus differentiation learning. The set of names will be referred to as a symbolic nonspatial stimulus set since its elements are not ordered along a spatial dimension. Care was taken to make the elements of the third and fourth stimulus sets adequately legible, i.e., to minimize problems of discriminability.

In order to determine the effect of different S-R pairings (given the stimulus sets and the response set) the first three sets of stimuli were assigned to the response set in three different ways, as shown in Fig. 2. The first method of S-R mating maximized agreement with population stereotypes. Maximum correspondence of S-R pairs is unequivocal for the spatial and for the symbolic two-dimensional codes. However, in the case of the one-dimensional spatial code there is no strong population stereotype as to which two-dimensional response corresponds to each of the elements in the row of stimulus lights. On the basis of data from a preliminary investigation the procedure adopted was to assign responses in a clockwise order letting the light at the left end of the row signify a response in the 7:30 o'clock direction.

The second method of S-R mating consisted of reversing the left-right relations in the maximum correspondence ensemble while retaining the relations in the vertical dimension. This condition will be referred to as mirrored correspondence among S-R pairs.

The third method of S-R mating was random assignment of stimuli to responses. Different sets of eight randomly-matched pairs were used with each *S*.

Since first names are not ordered along any spatial dimension they were also assigned to re-

sponses at random, the order being different for each *S*. Random assignment was the only method of mating studied with this stimulus set.

Light stimuli were presented to *S* by means of switches at *E*'s console. The other stimuli (numbers and names) were presented by opening a shutter behind a 1-in. aperture in the panel facing *S*. The selection of a stimulus to appear in the aperture, and the opening of the shutter, was accomplished by remote control from *E*'s console. Reaction time was measured from either the onset of the light or the opening of the shutter.

The stimulus panel was located 28 in. away from and 15° below *S*'s eyes, and perpendicular to his line of sight. The response panel was located on the tilted surface of a table in a comfortable position in front of *S*. A movement of the stylus toward *S* was the appropriate response to a stimulus in the 6 o'clock position. Such a movement agrees with the population stereotype for this situation.

Subjects.—One hundred volunteer college students served as *S*s. They were scheduled in pairs of the same sex, and were assigned to one of the ten experimental groups by means of a table of random numbers, the only restrictions being that each group contain ten *S*s and the same proportion of men and women.

Procedure.—The nature of the task was explained briefly. Each *S* was then allowed to study a diagram indicating how stimuli had been assigned to responses in his learning task. After 1 min. the diagram was removed and *S* was asked to indicate the appropriate responses to each stimulus by drawing arrows on an answer sheet. One-minute study periods interspersed with test periods were continued until each *S* met a criterion of two successive errorless performances. Learning of the spatial and of the symbolic two-dimensional codes, with either optimum or mirrored S-R correspondence, seldom required extra sessions beyond the two necessary to meet the criterion. However, the randomly-assigned spatial one- and two-dimensional codes required 70% and 75% more trials, respectively, than the minimum number possible. The randomly-assigned symbolic codes (numbers and names) required only 30% and 15% more trials, respectively, than the minimum necessary to meet the criterion.

After the familiarization procedure was completed, appropriate instructions were read to each pair of *S*s. The instructions for the spatial two-dimensional stimuli are typical and were as follows:

"Hold the stylus in your preferred hand. When I say 'ready' place the point on this button and hold it there. Two to four seconds after I say 'ready' one of these eight lights will come on. Slide the stylus as quickly as you can along one of the paths in the direction indicated by the light. If you move to the wrong position the light will remain on, and you should correct your mistake as quickly as possible. When you have completed a correct response I shall turn out the stimulus light."

The *S*s were further instructed to emphasize accuracy rather than speed and to limit their errors to approximately 1 in 16 responses.

The *E* sat where he could observe *S*'s hand movements and recorded the direction of the initial movement and of any subsequent correction movements. Errors are defined as movements which were sufficiently large in amplitude to enter a wrong pathway (see Fig. 1). Since *S*s were trying to complete their responses as quickly as possible nearly all movements that were started in the wrong direction were of relatively large amplitude and were easily detected.

Each *S* served for two 1-hr. sessions on different days and was tested under one of ten experimental conditions. During each session he received 64 trials, 8 to each stimulus, or a total of 128 trials during the two days. Each *S* also observed his partner for an equal number of additional trials. Trials were given in blocks of 16 responses, the pair of *S*s alternating between observing and responding. At the end of each of these sequences *S* was told the number of errors and the accumulated reaction time for the 16 responses. All *S*s appeared to be well motivated.

Note that in all experimental groups amount of information per stimulus was held constant at 3 bits per stimulus event.

Results

Reaction time and errors.—The means and *SD*'s of the pooled reaction time measures and the mean and median percentage of errors for each group are given in Table 1. Results for each of the two days as well as the combined data are included.

On both days the fastest mean reaction time and the fewest errors were recorded for the group using spatial two-dimensional stimuli (lights arranged in a circular pattern) mated with responses in such a way as to provide maximum agreement with population stereotypes. This is the only coding condition that maximizes both of the factors hypothesized to determine the degree of compatibility. The most errors, and the second slow-

TABLE 1

SUMMARY OF REACTION TIME AND ERROR DATA
(Groups contained 10 *Ss*, each of whom made 64 responses per day for two days)

Condition	Reaction Time (Seconds)						Errors (Per Cent)					
	Day 1		Day 2		Combined		Day 1		Day 2		Combined	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	Mdn.	Mean	Mdn.	Mean	Mdn.
Spatial 2-dim.												
Maximum corresp.	.421	.052	.354	.033	.387	.035	2.2	.8	1.6	.8	1.9	.8
Mirrored corresp.	.572	.085	.510	.081	.541	.069	5.5	5.6	3.3	2.8	4.4	3.9
Random corresp.	1.263	.186	.960	.120	1.111	.146	20.5	20.3	10.3	10.9	15.1	13.7
Symbolic 2-dim.												
Maximum corresp.	.707	.118	.644	.158	.675	.134	5.9	5.5	4.2	4.2	5.0	4.8
Mirrored corresp.	.844	.109	.709	.111	.777	.106	10.0	9.4	4.4	4.7	7.2	6.6
Random corresp.	.996	.337	.774	.121	.885	.218	12.1	10.9	7.8	7.8	10.0	8.6
Spatial 1-dim.												
Maximum corresp.	.853	.127	.733	.114	.793	.108	15.0	11.7	9.8	8.6	12.4	10.2
Mirrored corresp.	.928	.168	.748	.118	.838	.137	19.5	18.7	11.4	7.8	15.5	13.3
Random corresp.	1.339	.212	1.146	.311	1.242	.236	18.9	20.3	8.6	7.8	13.8	9.7
Symbolic nonspatial												
Random corresp.	.900	.150	.743	.134	.821	.129	14.2	12.5	9.7	10.1	11.9	9.4

est time, were recorded for the same two-dimensional spatial stimulus set, when random S-R pairing was employed. With this particular stimulus set mean reaction time was approximately three times as long and errors were approximately eight times as frequent for random as for maximum S-R pairing.

The symbolic two-dimensional stimulus code (clock numbers) with maximum S-R correspondence resulted in much slower reaction time and more errors than did the two-dimensional spatial code with either optimum or mirrored S-R correspondence. However, these relations were reversed for random S-R pairing of the latter stimulus set.

Improvement in reaction time from the first to the second day was small, but improvement in accuracy was rather large. For any given block of training trials, however, these two criteria agree closely in the rankings given to the ten coding conditions. For example, when all data are pooled over the two days the product-moment correlation between mean error scores and mean reaction time for the ten coding conditions was .84, which is significant at the $p < .01$ level of confidence. Since the reaction time measures are the more reliable, an analysis of variance was carried out on these scores for nine of the experimental conditions. The tenth condition, the symbolic nonspatial stimulus set with random S-R pairing, was not included in this analysis since it is not possible to achieve maximum or mirrored S-R correspondence with this set.

For the analysis of variance the conventional logarithmic transformation of the reaction time scores was made after the scores had been properly coded to avoid negative logarithms. Bartlett's test on the log scores supported the assumption of homogeneity. The results of the analysis are summarized in Table 2. The design employed for the analysis has been discussed elsewhere (**10**). Compari-

sons between experimental conditions are based on different groups of *Ss*, but the test for learning is based on the same *Ss*.

Although the effect of days was highly significant ($p < .01$) the absolute amount of improvement in speed from Day 1 to Day 2 was relatively small and there was no significant interaction between either experimental variable and days. The results therefore support the conclusion that compatibility effects do not change significantly with practice over two days. It should be noted that any transient learning effects that might have occurred during the first few trials on the first day are obscured by combining all the data for a given day.

An important finding with respect to the two main experimental variables was their highly significant interaction. This interaction apparently arises from the fact that the degree of compatibility of some S-R ensembles (those employing stimuli exhibiting a weak population stereotype) is affected only slightly by the method of assigning stimuli to responses, whereas the degree of compatibility characterizing other S-R ensembles (those exhibiting a strong population stereotype) is markedly affected by the method of forming S-R pairs. The latter, of course, are the stimulus coding sets for which maximum positive transfer or maximum interference effects would be expected. When tested using the interaction term, the method of pairing stimuli and responses was significant only at the .05 level of confidence. Since neither of the experimental conditions was introduced as a random variable, the latter finding cannot be generalized without reservation. In other words, random pairing of stimuli and responses in itself does not necessarily lead to poor performance—the effect depends on the strength of the population stereotype for a given stimulus coding set.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF LOG REACTION TIME DATA FOR NINE OF THE EXPERIMENTAL CONDITIONS

Source	Sum of Squares	*df*	*F*†
Stimulus sets (C)	.981	2	2.75
S-R pairings (P)	1.988	2	5.57*
C×P Interaction	.714	4	16.38**
Residual (net between *Ss*)	.882	81	
Learning (L)	.267	1	133.50**
C×L Interaction	.001	2	—
P×L Interaction	.009	2	2.25
C×P×L Interaction	.011	4	1.40
Residual (net within *Ss*)	.078	81	
Total		179	

† The main effects (C and P) are tested using the C × P interaction term; the C × P interaction is tested using the residual (net between *Ss*) term.
* $.05 > p > .01$.
** $p < .01$.

The preceding results are based on the mean reaction time for all eight elements in an S-R ensemble pooled. Mean reaction times were also computed for each S-R pair in each ensemble. There was somewhat less variability among the eight responses of the more compatible S-R ensembles than among those of the less compatible ones. The results agree with those of Garvey and Knowles (7). Among the ten experimental groups variability was greatest among the S-R pairs forming the one-dimensional spatial code. Reaction time for the two end elements of this set was consistently faster than for the centrally located elements under all three methods of S-R pairing. In terms of the present theoretical formulation this finding suggests that additional information transformation processes were involved in responding to the center lights.

Information analysis.—An analysis was made of the average amount of information transmitted per stimulus

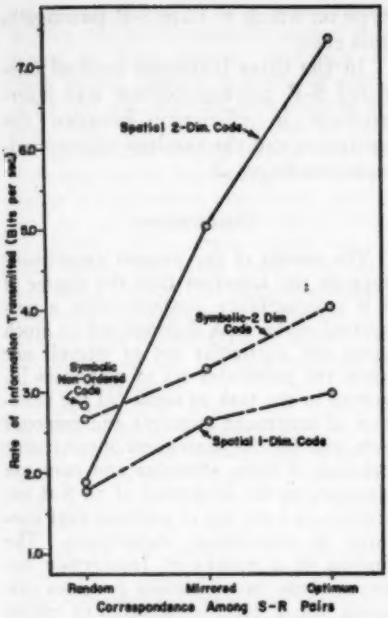


FIG. 3. Average rate of gain of information in bits per second for the ten experimental conditions, calculated from discrete reaction time and error data, pooled for all *S*s in a group

by each of the ten experimental groups. The computations were made from frequency tables in which the errors for all *S*s in a particular group were pooled. These tables were set up in a conventional manner for optimum and for mirrored pairings of the elements within S-R ensembles. However, in the case of the four groups using random S-R pairings (groups in which pairings were different for each *S*) the pooling of data obviously could not be made on the basis of identical stimuli. Instead, error frequencies were tabulated on the basis of the response that was designated by each stimulus. Thus, if one *S* had been trained to give a 12 o'clock response to the word *Vic*, and another *S* had been trained to give the same response to the word *Ben*, then these two stimulus words were considered equivalent.

After the pooled information transmitted per stimulus had been computed for each of the ten groups, these estimates were divided by the mean reaction time per response for that experimental group to obtain an estimate of the average rate of gain of information per stimulus for the different types of S-R ensembles. The results, shown graphically in Fig. 3, illustrate the magnitude of the interaction between coding dimension and method of pairing stimuli and responses. It should be noted that the present procedure may give a somewhat different estimate of the rate of information processing than that obtained from a serial task in which one stimulus follows another in rapid succession. However, computation of the rate of gain of information (8) is a useful way of combining speed and error data obtained in discrete response tasks.[2]

The lowest rate of gain of information, about 1.8 bits per second, occurred for the two sets of spatial stimuli with random S-R pairing. Random S-R pairings within the two symbolic stimulus sets (clock numbers and first names) gave considerably higher rate of gain of information than did random pairings within the two spatial stimulus sets. Unfortunately, no theoretical sampling distribution for these pooled estimates of information processing is available for testing directly the significance of the latter differences. However, an indirect test was made. Individual *t* tests were computed between the mean reaction times for the four groups. The results are shown in Table 3. The means for the two spatial (pic-

[2] Following Hick's convention the term "rate of gain of information" is used when speed and error data from discrete responses are combined; the term "rate of transmission of information" is reserved for the serial or continuous response case.

torial) codes did not differ significantly from each other, and the means for the two symbolic codes did not differ significantly from each other. However, all comparisons between a symbolic and a spatial code were significantly different in favor of the former. The proportion of errors was also smaller for each of the symbolic than for either of the spatial codes. Since information rate is a function of these two variables (errors and time), it is reasonable to assume that the variation in rate of gain of information among the four groups is also significant. It can be stated with some confidence, therefore, that the interaction between coding sets and methods of forming S-R pairs, which appears as a reversal in the rate of gain of information for the pictorial and the symbolic two-dimensional codes when the method of S-R pairing is shifted from optimum to random, is a significant effect.

It is of interest to note that the differences between the symbolic and spatial codes, with random S-R pairing, were all in the same direction for the 1-min. preliminary learning trials as for the experiment proper.

The spatial one-dimensional code consistently led to a low rate of information transfer. As noted earlier, there is no strong population stereotype on which to base S-R pairings in this case.

In the three instances studied mirrored S-R correspondence was intermediate in efficiency between the optimum and the random correspondence condition.

TABLE 3

SUMMARY OF t TESTS BETWEEN THE MEAN LOG REACTION TIMES FOR THE S-R ENSEMBLES COMPOSED OF RANDOMLY MATED S-R PAIRS

Condition	Symbolic	Spatial	
	Nonspatial	2-Dim.	1-Dim.
Symbolic			
2-dimensional	.58	2.83*	3.61**
Nonspatial		4.60**	5.15**
Spatial			
2-dimensional			1.37

* $.01 < p < .05$, 18 *df*.
** $p < .01$, 18 *df*.

DISCUSSION

The results of the present experiment support the assertion that the degree of S-R compatibility characterizing a perceptual-motor task depends not so much upon the particular set of stimuli nor upon the particular set of responses involved in the task as upon (*a*) the selection of congruent stimulus and response sets, and (*b*) the generation of congruent pairings of these stimulus and response elements in the formation of an S-R ensemble, i.e., the use of pairings that conform to population stereotypes. The finding of a significant interaction between these two variables provides evidence for a third compatibility effect, namely, that the greater the spatial correspondence of stimulus and response sets, the more detrimental the effect of noncongruent S-R pairings.

In regard to compatibility effects that result from the selection of stimulus and response sets, it must be kept in mind that the present experiment employed only one set of responses. It could therefore be argued that the variations in performance found for different stimulus sets could represent stimulus differences per se. Since the stimuli were sufficiently above threshold to be adequately visible and legible, however, such an argument is hardly tenable in view of the results of the previous study where it was shown that the best set of stimuli for one set of responses could become the worst set of stimuli with another set of responses. As an illustration, consider the row of eight lights, which was the worst of the sets of stimuli used in the present study. It can be argued that if the task had been to point to the light that was on, to push a key located under the light, or to move a joy stick to the right by an amount proportional to the distance from some reference point to the light, then this particular

set of stimuli would have been superior to, say, the set of symbolic stimuli composed of first names. Although the data of the present experiment admittedly do not directly substantiate this argument, the authors prefer to make the theoretical interpretation that the difficulty met in responding to the row of lights in the present experiment was not due to lack of discriminability among the lights, in the conventional usage of that term (the lights were 1 in. apart and easily visible), but to some additional information transformation process (such as counting) that intervened between the occurrence of the stimulus and the appropriate response.

Compatibility effects are relatively large in magnitude when compared with either short-term learning effects or with the effects of variations in the amount of information per stimulus. For example, depending on the method of S-R pairing employed, the amount of information that was gained per second with the two-dimensional spatial code varied from a value (1.8 bits per second) which is slightly less than required to designate 1 from among 4 equally likely alternatives per second to a value (7.3 bits per second) equivalent to the designation of 1 from among about 160 alternatives per second. The differences attributable to changes in S-R pairing in the present experiment are comparable in magnitude to those found by Morin and Grant (13) when they introduced slight departures from direct spatial correspondence in assigning a row of stimulus lights to a row of response keys.

The authors would like to suggest that the concept of S-R compatibility might well replace such older concepts as meaningfulness and belongingness in specifying the conditions of motor learning. It may also prove useful in sharpening the definition of similarity (6). In studies of transfer effects, for example, it may be important to specify that the task to be learned is one for which there is no strong population stereotype, as was true of the task studied by Duncan (2). It may be even more important to specify that transfer is from a relatively compatible to a relatively incompatible task (as was true of the learning tasks studied by Morin and Grant [13] and by Lewis [11]), from one incompatible task to another, etc.

Compatibility effects also should be considered in measuring individual differences in perceptual-motor ability. Mitchell and Vince, for example, have suggested that "when performance is affected by a nonpreferred relationship, the task becomes one where the cognitive element plays a large part . . ." (12, p. 34). Although some widely-used psychomotor tests, such as the Two-Hand Coordination Test, employ incompatible S-R relations (14), the relation of this task characteristic to the abilities measured by such tests has not been stressed.

There are many obvious applications of the concept of compatibility to the design of tasks for most efficient learning and performance. In the present study, for example, the marked superiority of an optimal pictorial code over an optimal symbolic code for use with a set of spatial responses is of considerable practical and theoretical significance for human engineering.

Previous studies (4, 13) have shown that the efficiency of information processing varies among different S-R ensembles even after extensive training. The present experiment adds to this evidence the findings that results were stable over two days, and that a particular two-dimensional spatial stimulus was superior to a particular symbolic stimulus. The latter result is important because the difference occurs in spite of many years of experience and a strong population stereotype in interpreting the symbolic stimulus in spatial terms.

Summary

An experiment was conducted to test the hypothesis that S-R compatibility is maximum when the pairings of stimulus and response elements in the formation of an S-R ensemble insure maximum agreement with population stereotypes. The experiment permitted a further test of the previously examined hypothesis that maximum S-R compatibility requires correspondence of stimulus sets and response sets in respect to the dimensions along which stimulus and response categories are selected, and also an evaluation of the interaction of the choice of stimulus sets with the method of S-R pairing.

Compatibility effects are conceived as arising

from an intervening information transformation process which is indicated by the statistical interactions of stimulus sets, response sets, and S-R mating procedures. The hypothesis was tested by analyzing the reaction times and errors of *Ss* in a series of different experimental situations.

Ten groups, each containing ten randomly-assigned *Ss*, were studied. Two spatial stimulus sets, a circle and a row of lights, and two symbolic stimulus sets, clock numbers and first names, were employed in forming S-R ensembles. Each of the stimulus sets contained eight alternatives. A single response set, composed of eight directional motor responses, was used for all groups. Stimuli were paired with responses in three ways, so as to provide maximum, mirrored, and random S-R correspondence (only random assignment was used with the set of first names).

The chief finding with regard to the two main experimental variables (choice of stimulus coding set and mating of S-R pairs) was their highly significant interaction. This finding and subsequent *t* tests support the original hypothesis. With either optimum or mirrored mating of S-R pairs the spatial two-dimensional ensemble was superior to all other groups. Performance was generally poor with random S-R mating, but significantly better performance was achieved in this case with the two symbolic coding sets than with either of the two spatial coding sets.

All groups were tested over two days. Improvement in reaction time between days was highly significant but of relatively small absolute magnitude, and there was no significant interaction between days and any of the experimental (compatibility) effects. Considered in relation to the performance differences between different coding procedures, and in the light of the finding from the previous study in this series, these results indicate that compatibility effects in perceptual-motor tasks are relatively large in comparison with effects produced by short-term learning or by changes in the number of alternatives (amount of information) relevant to each successive choice. The effects also appear to be relatively permanent.

Implications of the concept of compatibility for studies of transfer of training and individual differences are discussed.

References

1. Andreas, B. G., & Weiss, B. *Review of research on perceptual-motor performance under varied display-control relationships.* Rochester, N. Y.: Univ. of Rochester, 1954. (Sci. Rep. No. 2, Contract No. AF 30 [602]-200.)
2. Duncan, C. P. Transfer in motor learning as a function of degree of first-task learning and inter-task similarity. *J. exp. Psychol.*, 1953, **45**, 1–11.
3. Fitts, P. M. Engineering psychology and equipment design. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1287–1340.
4. Fitts, P. M., & Seeger, C. M. S-R compatibility: spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, 1953, **46**, 199–210.
5. Fitts, P. M., & Simon, C. W. Some relations between stimulus patterns and performance in a continuous dual-pursuit task. *J. exp. Psychol.*, 1952, **43**, 428–436.
6. Gagné, R. M., Baker, K. E., & Foster, H. On the relation between similarity and transfer of training in the learning of discriminative motor tasks. *Psychol. Rev.*, 1950, **57**, 67–79.
7. Garvey, W. D., & Knowles, W. B. Response time patterns associated with various display-control relationships. *J. exp. Psychol.*, 1954, **47**, 315–322.
8. Hick, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–27.
9. Knowles, W. B., Garvey, W. D., & Newlin, E. P. The effect of "speed" and "load" on display-control relationships. *J. exp. Psychol.*, 1953, **46**, 65–76.
10. Kogan, L. S. Analysis of variance—repeated measurements. *Psychol. Bull.*, 1948, **45**, 131–143.
11. Lewis, D., McAllister, D. E., & Bechtoldt, H. P. Correlational analysis of the learning and relearning of four different tasks on the modified Mashburn apparatus. *Amer. J. Psychol.*, 1953, **36**, 83–109.
12. Mitchell, M. J. H., & Vince, M. The direction of movement of machine controls. *Quart. J. exp. Psychol.*, 1951, **3**, 24–35.
13. Morin, R. W., & Grant, D. A. Spatial stimulus-response correspondence. *USAF, WADC Tech. Rep.*, 1953, No. 53-292.
14. Norris, E. B., & Spragg, S. D. S. Performance on a following tracking task (Modified SAM Two-Hand Coordination Test) as a function of the plane of operation of the controls. *J. Psychol.*, 1953, **35**, 107–117.
15. Nystrom, C. O., & Grant, D. A. Performance on a key pressing task as a function of the angular correspondence between stimulus and response elements. *USAF, WADC Tech. Rep.*, 1954, No. 54-71.
16. Simon, C. W. Instrument-control configurations affecting performance in a compensatory pursuit task. *USAF, WADC Tech. Rep.*, 1952, No. 6015.

(Received for early publication July 31, 1954)

Journal of Experimental Psychology
Vol. 48, No. 6, 1954

AUTOKINETIC MOVEMENT OF VERY LARGE STIMULI [1]

WARD EDWARDS

Armament Systems Personnel Research Laboratory, Air Force Personnel and Training Research Center, Lowry Air Force Base

A recent paper (1) reported that large, bright stimuli often show autokinetic movement if they are carefully prevented from illuminating surrounding objects. With the methods used in that experiment, there was a systematic increase in latency before movement was seen and a systematic decrease in total amount of movement seen as size increased. That finding raises the question whether still larger stimuli might not show any autokinetic movement, or, alternatively, whether stimuli so large they fill half the visual field might still show such movement under the proper circumstances. This experiment is addressed to these questions.

METHOD

Apparatus.—The apparatus consisted of a matrix of dots, illuminated from behind. The dots consisted of light shining through holes, ¼ in. in diameter, in a screen. Behind the screen was a large ground-glass screen, and behind that was a bank of 24 matched photoflood bulbs. This arrangement produced the same color and brightness for all dots. The bulbs were operated at a very low voltage, so the dots were very dim and orange. There were 15 rows and 12 columns of dots. The distance from the midpoint of any dot to the midpoint of any other which adjoined it vertically or horizontally was 3 in. Thus the matrix was 42 in. long and 33 in. high.

Whenever the photofloods were turned on, a small electric motor was also turned on. This was not connected to anything; it merely served as a sound effect.

Subjects.—The *S*s were three psychologists and three unsophisticated *S*s who knew nothing about autokinetic movement. However, the unsophisticated *S*s were told in detail all about autokinetic movement before participating in the experiment. There were no systematic differences between the results for these two groups of *S*s, so they will be combined in all analyses reported here.

Procedure.—The *S* was taken into the dark room. While he was dark adapting he was told about autokinetic movement, told that any one of a wide variety of possible movements might occur, and asked to report all kinds of movement he saw, no matter how improbable they seemed. He was then seated 117 in. from the dot matrix. For the first five trials, only the outer rectangle of dots and a single dot in the center were presented. The rest were covered by a mask. The *S* was instructed to fixate the single dot. As soon as *S* saw any kind of movement of any part or all of the stimulus, he pressed a button which started a timer and stopped a clock which recorded his latency. Thirty seconds after *S* pressed his button, the lights and motor were turned off, and *S* reported what he had seen. He reported all distances as percentages of the vertical dimension of the matrix; this gave him a stable unit with which to measure length. After five trials, the mask was removed and the whole matrix was presented for five trials. The *S* was instructed to choose a central dot as a fixation point. Then the *S* was moved to a chair 46 in. from the dot matrix, and the whole process was repeated.

There are several ways of measuring the visual angles of the stimuli used in this experiment. Table 1 shows these various visual angles.

RESULTS

All *S*s reported movement on all trials. The reason presumably is that each stimulus was left on until *S* pressed his button, indicating he had seen movement. The longest latency was 151 sec.—2½ min. However, the mean latency was only 28.97 sec. An analysis of variance of the latency data showed that *S*s differed from each other very significantly, and the difference between the near and the

[1] This experiment was done under Contract N5-ori-166, Task Order 1, between the Office of Naval Research and The Johns Hopkins University. This is Report No. 166-I-190, Project Designation No. NR 145-089, under that contract. The research was done while the author was at The Johns Hopkins University.

TABLE 1

VISUAL ANGLES SUBTENDED BY THE STIMULI AT TWO VIEWING DISTANCES

Dimension of Stimulus	Viewing Distance	
	46 In.	117 In.
Horizontal	49°05′	20°21′
Vertical	39°28′	16°04′
Diagonal	60°16′	25°43′

far positions of viewing was significant, with the near position, surprisingly, having the lower latency. The difference between the dot-and-rectangle pattern and the complete matrix fell just short of significance, with the former having the lower latency. There was a significant triple interaction between *S*s, pattern, and distance. This triple interaction might be attributed to the effect of sequence of presentation on some *S*s, since the same sequence was used for all *S*s. However, examination of the data makes this seem unlikely, since the effect does not consist of any orderly change (e.g., a decrease in latency) with successive experimental arrangements. Instead, the most probable cause of this significant triple interaction is that some *S*s had unusually long latencies on the first group of presentations of the complete matrix.

All kinds of movements were reported, and usually more than one kind and direction of movement was seen in any single trial (this is the natural result to be expected when a complex pattern is presented for as long as 30 sec. after movement begins). The *S*s looking at the dot-and-rectangle pattern most often reported movement of the whole pattern up and to the left or right, but some saw movement only of the dot, some saw movement only of the rectangle, and some reported that the two moved independently of one another. A very common report for both patterns was that the whole pattern rotated back and forth around its center. On the complete matrix, most *S*s saw movement of the whole pattern. One *S* saw the individual dots spinning. Another saw oscillatory movement in the third dimension. Usually, these minor movements would be combined with lateral or vertical movements of the whole pattern. A number of times *S*s reported saccadic movements of small extent. The maximum movement reported was a 66-in. movement of the whole matrix.

A number of other explanations besides autokinetic movement can be thought of for some of these movements. Small rotations of the head might produce rotational movement. Eyestrain and fatigue might produce apparent size and brightness changes which might be perceived as three-dimensional oscillation. Saccadic movements might be caused by saccadic eye movements. These considerations raise the question whether or not any real autokinetic movement was actually seen. To settle this question, the data were examined with some very rigid criteria of autokinetic movement in mind. The movement, to be considered autokinetic for the purposes of this analysis, had to be of the whole pattern; it had to be up, down, left, or right; it had to continue in substantially the same direction without jerks or retracings for a fair distance; and it had to move a total distance of at least 20% of the vertical dimension of the matrix, i.e., at least 6.6 in. Six reports (out of 30) met these requirements at each viewing distance for the complete matrix; two reports at the far distance and one at the near distance met them for the dot-and-rectangle pattern. This difference between pat-

terns is significant at the .02 level. Each *S* met the requirements at least once. The reason for the fact that the complete matrix apparently worked better than the dot-and-rectangle pattern presumably is that so many reports of movement of only the central dot were made for that pattern. Had the requirements been applied to the central dot alone, they would have been met 19 times (out of 60).

Discussion

This experiment demonstrates the occurrence of autokinetic movement of very large patterns of dots, subtending up to 60° of visual angle. It also suggests that a filled, continuous pattern of dots tends to produce somewhat better autokinetic movement than does an outline rectangle of dots with a fixation dot in the middle, although if movement of the fixation dot alone is considered, this conclusion is no longer valid.

Why is the ordinary visual world not seen as autokinetically moving? This seems to be the most interesting question about autokinetic movement. This experiment suggests strongly that it is not simply the fact that the visual world is filled, rather than empty, which stabilizes it. The fact that the complete matrix seems to be a better autokinetic stimulus than the dot-and-rectangle pattern suggests that autokinetic movement is associated with a "monotonous" visual world, one which either has very few objects in it, or else has many objects but arranged in a very regular and meaningless way. Another way of putting it might be to say that autokinetic movement occurs with a highly redundant visual world and does not occur when the visual world is full of information. This principle, if taken seriously, would suggest that an irregularly arranged field of dots should show less movement than a regular one containing the same number of dots in about the same total extent. By extension, it would also suggest that meaningful visual stimuli should show less autokinetic movement than meaningless ones of about the same extent. These are experimental problems, and the experiments are easy to do.

Summary

The *S*s observed the autokinetic movement of very large patterns of dots. The maximum visual angle involved was about 60°. All *S*s saw movement on all trials. A wide variety of different kinds of movement was seen. When very rigid criteria for autokinetic movement were applied, all *S*s still saw autokinetic movement on one or more stimuli, and all stimuli were seen as autokinetically moving. It is concluded that visual patterns of any extent may be seen as autokinetically moving. It is suggested that perhaps the ordinary visual world is stable because it contains much more information than the visual worlds used in autokinetic movement experiments.

Reference

1. Edwards, W. Two- and three-dimensional autokinetic movement as a function of size and brightness of stimuli. *J. exp. Psychol.*, 1954, 48, 391–398.

(Received for early publication July 23, 1954)

Journal of Experimental Psychology
Vol. 48, No. 6, 1954

THE RECONSTRUCTION OF ABBREVIATED PRINTED MESSAGES[1]

ALPHONSE CHAPANIS[2]

The Johns Hopkins University

An important concept which psychologists have adopted from modern communication theory is that of redundancy. In general terms, redundancy refers to the predictability of the elements in a message by virtue of their statistical characteristics. Some of the redundancy in English text arises from the fact that letters, spaces, and punctuation marks are not used equally often. The "e," "t," "o," "i," and "n" are used commonly whereas the "z," "j," "q," "k," and "x" are relatively rare. Another source of redundancy is in the statistical dependencies between letters, spaces, and punctuation marks. For example, except for some trivial exceptions, the letter "q" is always followed by "u" in ordinary English text. Given the letters "th," the probability is high that the next letter is a vowel. A period is almost invariably followed by a space. And so on.

Although there are no precise estimates of the exact amount of redundancy in the English language, Shannon has estimated that it may be as high as 50%. One of the grounds on which he makes this statement is that he says it is possible to reconstruct English text when 50% of it has been deleted (**2**, p. 26). Although he does not elaborate on the conditions under which these results were obtained, the evidence consists of some trials with a few samples of English text which Shannon tried primarily on himself and his wife.[3]

In any case, Shannon's primary interest in redundancy is that of the theoretician; he wants to know the maximum value for the English language. There remains the practical question: How good is redundancy to the average person? How much does it help him in certain ambiguous situations? It is this latter aspect of the problem to which this paper is directed.

The purpose of the present experiment was to explore systematically some of the practical consequences of redundancy in English text. Specifically, I have attempted to discover how well a heterogeneous sample of *Ss* can reconstruct English text when various amounts of it have been deleted. A secondary aim of this experiment was to relate the ability to reconstruct such abbreviated passages to general intelligence and to familiarity with grammar, spelling, and sentence construction.

[1] This research was done under Contract N5-ori-166, Task Order 1, between the Office of Naval Research and The Johns Hopkins University. This is Report No. 166-I-185, Project Designation No. NR 145-089, under that contract.

[2] I am indebted to Mrs. Grace D. Bell for her meticulous supervision of this study. She took over the project when it was about one-fourth completed and was thereafter entirely responsible for the preparation of materials, testing of *Ss*, and (in part) scoring of the data. She also assisted capably in many of the initial statistical tabulations. I am also grateful to the Misses Marjorie Causey, Inge Grebe, Helene L. Kuhn, and Judith T. Parker who spent many tedious hours typing over 1,000 different passages which were so garbled as to constitute a stenographer's nightmare.

[3] Shannon, C. E. Personal communication.

Materials

Thirteen prose passages, varying widely in difficulty, were selected for this experiment. They are:

Passage 1

"The properties which give rise to the unusual effects of ultrasonic waves follow from principles which are common to all wave phenomena. In a given medium the wavelength of a wave is inversely proportional to the frequency, so a high frequency implies a short wavelength. ; Furthermore, the directional" (Laufer, A. R. Ultrasonics. *Physics Today*, 1950, **3**, 8–15.)

Passage 2

"This week the Big Three of the Council of Foreign Ministers, which is the governing body of the North Atlantic Treaty Organization, met in New York to speed the defense of Europe. And, as if to strengthen the hands of the ministers, there are reports from Washington that the Administration has decid" (Anonymous. The defense of Europe. *Life*, September 18, 1950, **29**, No. 12, p. 59.)

Passage 3

"Fascism sprang from an emergency and expressed the need of prompt and decisive action, ideologies being disparaged because they promote discussion, delay, and irresolution: pragmatism was, therefore, a welcome ally, for having boldly challenged the prestige and authority of the intellect. Fascism wa" (Perry, R. B. *The thought and character of William James*. Cambridge, Mass.: Harvard Univ. Press, 1948, p. 319).

Passage 4

"I marked him for white sugar," she confided. "He cain't he'p hisself. When I was carryin' him, we had nothin' but brown sugar. Willy—that was my husband—Willy was goin' to Orange Springs and I said to him, 'Willy, I've took the biggest notion for some white sugar. Fetch me back a half a pound, fo" (Rawlings, M. K. *South moon under*. New York: Bantam Books 1945, p. 195.)

Passage 5

"The fight fans hit the jackpot next Monday night at the Coliseum with a card of three ten-round co-features. Two of the skirmishes involve welterweights, the third brings in for consideration a pair of new heavyweight prospects. The young heavies listed here for the first time are Ted Calaman, of Ph" (Anonymous. New heavies here Monday. Baltimore: *The Evening Sun*, November 1, 1950, Vol. 82, No. 13, p. 59.)

Passage 6

"Republican leaders were reported today to have chosen Harold E. Stassen, 1948 presidential aspirant, to reply for their party to President Truman's Saturday night campaign speech. They turned to Stassen, President of the University of Pennsylvania, after Senator Arthur H. Vandenberg (R., Mich.) begg" (Anonymous. GOP is reported set on Stassen for radio talk. Baltimore: *The Evening Sun*, November 1, 1950, Vol. 82, No. 13, p. 1.)

Passage 7

"He was an easy sort of man. His blond hair was burnt with the sun and wind. His very blue eyes rarely lost their touch of laughter, of some profound and secret amusement that nudged his soul, day in, day out. He was strongly made, and his shoulders seemed, perhaps, a trifle broader than they really" (Markey, M. *Unhurrying chase*. Stamford, Conn.: Stamford House, 1946, p. 4.)

Passage 8

"The more adults you get to know, the more you will realize they are not alike. Each one is different, like children. Parents are the only ones that are important to babies, but after that you have to branch out a little. If you think they are all like your mother, for instance, you will get fooled." (Owsley, J. Grownups are funny people. *Parents' Magazine*, April, 1950, **25**, p. 36.)

Passage 9

"Line pie plate with pastry. Pare, core, and cut apples in eighths, put row around plate ½ inch from edge, and work towards center until plate is covered; then pile on remainder. Mix sugar, nutmeg, salt, lemon juice, and grated rind, and sprinkle over apples. Dot over with butter. Wet edges of unde" (Flesch, R. *The art of plain talk*. New York: Harper, 1946, p. 144.)

Passage 10

"It should also be noted that under the analytical treatment the ammonium sulfide metals may show change of valence, exhibit amphoteric properties, undergo hydrolysis, and form complexes with the ammonia ion. The hydrolysis of aluminum sulfide, which in systematic analysis leads to its precipitation" (Gordon, N. E. *Introductory college chemistry*. New York: World Book Co., 1926, p. 561.)

Passage 11

"It is theoretically not impossible to develop the statistics of the semantic and behavior languages to such a level that we may get a fair measure of the amount of information in each system. At any rate, we can show by general

considerations that phonetic language contains less over-all information" (Wiener, N. *The human use of human beings: cybernetics and society.* Boston: Houghton Mifflin, 1950, p. 91.)

Passage 12

"It is customary also for a factual writer to reveal, not to hide, his plan of organization. In this respect informative writing contrasts with essays and stories. Since the reader of a factual article reads to learn and to remember, and since an outline arranges facts into easily learned and remembe" (Guthrie, L. O. *Factual communication: a handbook of American English.* New York: Macmillan, 1949, p. 19.)

Passage 13

"The body of a nerve cell consists of a mass of granular cytoplasm surrounding a nucleus, which contains a well-defined nucleolus. Running through the cytoplasm and processes is an arrangement of fine fibrils called neurofibrils. They form a reticulum in the cell-body, are present in the dendrons, an" (Kimber, D. C., Gray, C. E., & Stackpole, C. E. *Textbook of anatomy and physiology.* New York: Macmillan, 1934, p. 153.)

Deletions were made from these passages in either random or regular patterns. For each deletion pattern, six amounts of text were deleted: 10, 20, 25, 33.3, 50, and 66.7%. Since there are 300 units in each of the 13 passages, the actual numbers of units deleted were 30, 60, 75, 100, 150, and 200. Letters, spaces, and punctuation marks were counted as units to be deleted. Every regular deletion pattern was also applied in all of its possible positions with respect to the beginning of the passage. For example, two of the 10% regular deletion patterns for each passage were constructed by deleting the first, eleventh, twenty-first, etc., letters (or spaces, or punctuation marks); two were made by deleting the second, twelfth, twenty-second, etc.; two were made by deleting the third, thirteenth, twenty-third, etc.; and so on. This same procedure of varying the starting position was applied to all the other amounts of deletion.

The random deletion patterns were taken from a table of random numbers and were so selected that they gave the exact percentages required for deletion. Every random pattern was different from every other one.

The abbreviated passages were made up into test booklets. Each test booklet contained 13 passages: one control passage (from which no material had been deleted) plus one sample of each of the 12 combinations of deletion pattern and amount of deletion. Each passage appeared only once in every test booklet.

Method

Procedure.—Each test booklet had a page of instructions mimeographed on the front of it, and the simplest way to describe the procedures is to quote part of these instructions:

"In this booklet you will find 13 paragraphs from which letters, spaces, and punctuation marks have been left out. In some cases not very much has been left out; in other cases quite a bit has been left out; and in still other cases a great deal has been left out.

"We want you to rewrite every paragraph. Guess what each paragraph said in its original form, and fill in the material that was left out. Fill in as much as you possibly can, even if you cannot get entire words or sentences. If you can't make any sense out of something, copy what you see. If you find any sentences or paragraphs that seem complete, copy them as they are.

"There are certain rules of the game that you ought to know about:

"1. All the original passages began with a new sentence, but some end in the middle of a sentence—or the middle of a word. Go ahead and try to guess as many additional words as you can beyond the end of the message.

"2. A space always means the end of a word.

"3. Since spaces have sometimes been left out, several words may be run together. In other words, the end of a word is not always followed by a space.

"4. The end of a line always comes at the end of a word.

"5. Sometimes you will find more than one space. This means that one or more words have been left out. That is why two or more spaces come together. In order not to confuse you, however, there is only one space at the end of any line.

"On the next two pages are two sample paragraphs which have been written out. Please study them carefully. If you have any questions about the rules or how to write out the paragraphs, be sure to ask for an explanation."

These instructions were read aloud by *E* to each *S* as *S* followed the text himself. After this the following additional instructions were read to *S*. (These additional instructions were evolved after the first few *S*s were tested, and were designed to answer some recurrent questions.)

"1. All the material given in the passages is correct. We have not used any wrong letters, nor have we reversed the order of the letters. We have not used capitals, etc. where they do not belong. All we have done is to delete certain amounts of the material.

"2. If you get stuck on a passage, you can go on to the next one and come back to it later.

"3. Each passage is on a separate page.

"4. You can ask any questions any time you wish."

Following the instructions came two mimeographed sample passages. The *E* went over the practice problems with each *S*, pointing out examples of the various rules. The *S*s were repeatedly told that they should take all the time they needed to do the best possible job. In actual practice *S*s took from about 3 to 16 hr. to complete the booklet and usually distributed their efforts over several days. They were not allowed to use dictionaries, word lists, grammars, or any other aids while taking these tests.

In addition to the test booklet every *S* took the Cooperative English Test, Form Y, Test No. 331-84-1, and the Wonderlic Personnel Test, Form A. These tests were administered in the standard way.

Scoring the data.—Five different measures of performance were obtained for every sample passage:

1. The total number of letters, spaces, and punctuation marks inserted.

2. The total number of letters, spaces, and punctuation marks correctly inserted. In general, scoring was liberal since credit was given for correctly inserted letters even though it was apparent that *S* missed the sense of the passage. For example, the word "through" once appeared as "thgh" and *S* guessed it to be "though." He received credit for the "o" and "u" even though the word was not correct. As another illustration, the words "to such" appeared in one of the passages as "touc." The *S* guessed this to be "touch" and was given credit for the "h" although he obviously did not get the correct sense of the phrase.

3. The number of items incorrectly copied or omitted in copying from the text which was supplied.

4. The number of items attempted beyond the end of the passage.

5. The number of items correct beyond the end of the passage. The scoring here was strict and literal. Letters, spaces, or punctuation marks were counted only if they followed the original text exactly, and they were counted only up to the first error. This score is not contained in the second score discussed above.

All scoring was done by two persons—the author and Mrs. Grace D. Bell. There were relatively few instances of ambiguities which caused trouble in scoring. Whenever ambiguities were discovered, the two scorers went over the passage together until they arrived at unanimity of opinion. Although a reliability figure was not computed, it would undoubtedly have been very high. The scoring was tedious, but, contrary to our expectations, not difficult.

Subjects.—The *S*s in this experiment were 91 persons deliberately selected to represent a wide range of talent as regards general intellectual ability and general familiarity with the statistical structure of the English language. Five *S*s were in a special category—they were naval officers, code experts, working in the Bureau of Naval Intelligence.[4] Nine *S*s were naval enlisted men, high school graduates but with no specialized training. There were 23 high school students, 47 undergraduate and graduate college students, and 7 other persons (housewife, clerk, etc.). Seventy-two of the *S*s were male, 19 female. The ages of the *S*s ranged from 17 to 44 yr. with a median of 19. Most of the *S*s ranged from 17 to 31 yr. of age; the code experts had an average age of 34 yr. and they formed the oldest subgroup in the sample.

The distribution of scores on the Wonderlic test shows this group of *S*s to be above the average of the sample on which the test norms for all educational levels are based. The lowest score, 11, corresponds to the 9th percentile; the highest score, 43, to the 99th percentile; and the average score, 27.8, to about the 70th percentile. The *SD* of the distribution is 7.2. The scores on the Cooperative English Test range from 33 to 79 with an average of 58.0 (*SD* = 10.7). These correspond to the 1st, 99th, and 54th percentiles, respectively, for freshmen entering the typical liberal arts college. In both distributions, the five naval code experts were well above the average.

Results

Amount of material inserted.—The first question of interest is, "How much material did *S*s supply for these test passages?" Figure 1 summarizes the data on this point with no distinction being made between incorrect and correct insertions. (Figure 1, however, does not include items guessed beyond the end of the passage.) For passages containing up to 25% deletions, there is an almost linear increase in the number of items supplied. When a third or more of the text has

[4] I am grateful to Captain J. A. Morrison, Office of the Chief of Naval Operations, for his enthusiastic cooperation in enlisting the services of these specialists.

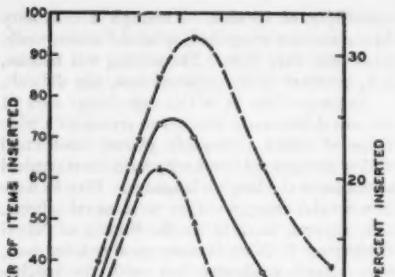


FIG. 1. Number of items (both correct and incorrect) inserted into the test passages by the 91 *Ss* in this study. The three points above each abscissa value except 0 are calculated from distributions containing 182 cases (two passages —one with regular, the other with random deletions—for each of the 91 *Ss*); above the abscissa value 0, the *N* is 91. The solid curve joins the medians of the distributions; it is not the curve of the median *S*. The scale on the right expresses the data as percentages of the original, uncut length of the passage.

been deleted, the total amount of material supplied decreases with increasing amounts of deletion until for passages with 67% deleted the median *S* attempts to supply very few items. Note, incidentally, that a considerable number of *Ss* inserted material into the control passages, i.e., passages from which no items had been deleted.

The dashed lines in Fig. 1 define the interquartile range and they show a marked increase in the variability as more and more text is deleted. When small amounts of text have been deleted, *Ss* supply just about the right amount of material. When greater amounts of text are deleted and the passages become more difficult, the variability increases due largely to individual differences between *Ss*. In general, *Ss* who scored low on the two standardized tests supplied very few items for the difficult passages. The *Ss* who scored high on these tests tended to be much more fluent and productive of items on the difficult passages. Individual differences and their correlates are discussed more fully later.

The data in Fig. 1 can also be looked at in terms of the fraction of the deleted material which is supplied by *S*. When 10, 20, or 25% of the text has been deleted, *Ss* supply about the right amount of material. For amounts of deletion greater than 25%, however, too little material is supplied. For example, when half of the text has been deleted, the median amount of material supplied is less than 25% of that required (37 items supplied as compared with 150 deleted).

Of the total amount of material in-

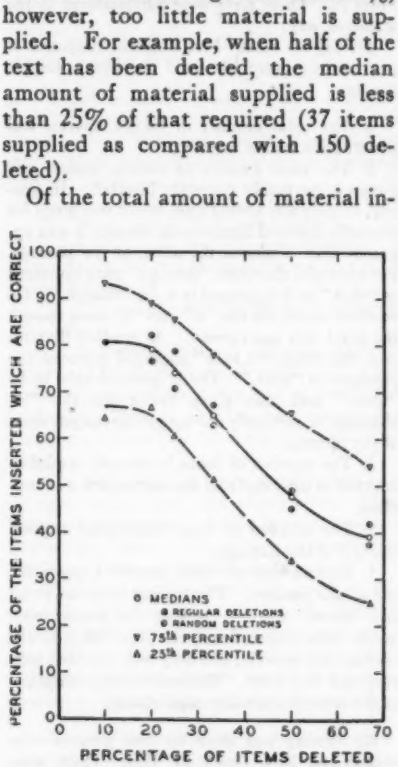


FIG. 2. Percentages of items correct of the total number inserted. Medians are shown separately for the regular and random deletion patterns, and for the combined distribution of 182 cases.

serted, not very much of it was correct (see Fig. 2). When only 10% of the text had been deleted, scarcely 80% of the material inserted was correct. As the amount of deleted text increases, this percentage decreases markedly. When half of the text had been deleted, less than half of the material supplied was correct.

There is a fair amount of variability in these percentages as is evidenced by the spread between the 25th and 75th percentiles. The upper limit of the range of percentages is 100 throughout, but this has little significance for the high amounts of deletion because it means only that a few *Ss* guessed a few items (see Fig. 1) and that all of their guesses were correct. For texts from which two-thirds of the material had been deleted, most *Ss* typically guessed at the spaces between words. This is perhaps the easiest kind of item to supply

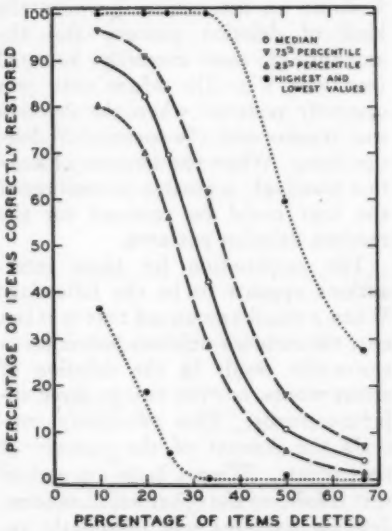


FIG. 3. Percentages of the deleted items which were correctly restored. Data are for 91 *Ss* and for regular and random deletion patterns combined so that the distribution above each abscissa value has 182 cases.

because capital letters occurring in the middle of long strings of letters frequently mean the beginning of a sentence, and so should be preceded by a space and period. A few *Ss* were able to supply 2, 3, or 4 such items correctly and so scored 100% on this measure. The lower limit of the range (0%) has little significance for similar reasons.

Amount of material correctly restored. —Whereas Fig. 2 shows the number of correct items as a percentage of the total number of items inserted, Fig. 3 shows the number of correct items as a percentage of the number of items deleted. In other words, Fig. 3 shows the percentage of the deleted material which was correctly restored. When 10% of the original text has been deleted, *Ss* are able to restore about 90% of it. Ability to do this falls off rapidly, however. When about 30% of the text has been deleted, just about half of it (by interpolation) can be correctly restored. When 50% or more of the text has been deleted, the median *S* is able to restore just about 10% of the deleted content. Most of the data cluster closely around the median curve, but the range of values is very large. The poorest *S* restored scarcely a third of the material when only 10% of the original text had been deleted. On the other hand, a few *Ss* were able to restore all of the material when as much as one-third of the text had been deleted. Two of the factors contributing to this wide scatter are (*a*) individual differences among *Ss*, and (*b*) individual differences among passages. Both of these will be discussed in more detail later.

The difficulty of the task we presented to our *Ss* can be illustrated by a count of the number of perfect restorations. In all, the 91 *Ss* saw a total of 1,092 passages from which material had been deleted in some

amount or other. Only 19 of these passages were perfectly restored; that is, all of the missing items were correctly replaced and no extra items were inserted. Thirteen of these perfect restorations were made with regular deletion patterns, six with random ones. Fifteen *Ss* each contributed one perfect restoration; two *Ss* each contributed two. In addition to the 19 perfect restorations, an additional 16 passages had all of their missing items restored, but extra items (not in the original text) were also supplied.

Regular versus random deletion patterns.—There are some small but consistent differences between the data for regular versus random deletion patterns. The differences show up as statistically significant interactions which are best illustrated graphically. Figure 4 shows that when the deletion pattern was regular, *Ss* supplied more material for passages from which a third or less material had been deleted. For passages from which a half or more of the text had been deleted, *Ss* supplied more material when the deletion pattern was random. Figure

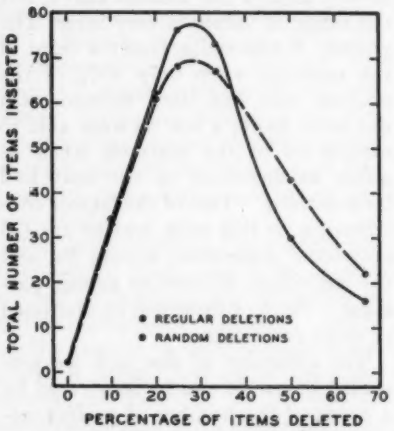


FIG. 4. Median numbers of items inserted for the regular and random deletion patterns separately. Each point is based on an *N* of 91.

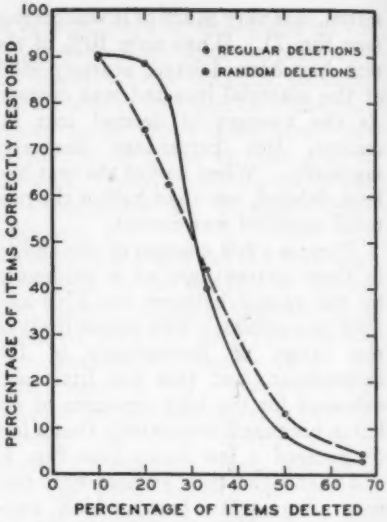


FIG. 5. Median percentages of items correctly restored for the regular and random patterns separately. Each point is based on an *N* of 91.

5 shows an interaction between the kind of deletion pattern and the amount of text correctly restored (see also Fig. 2). More text was correctly restored when the deletion was regular and the amount of deletion low. When the amount of deletion was high, a greater percentage of the text could be restored for the random deletion patterns.

The explanation for these interactions appears to be the following: When a small amount of text is taken out, the random deletion patterns occasionally result in the deletion of entire words, or even two or three adjacent words. This effectively conceals the context of the passage at that point. When a large amount of text is lost, on the other hand, random deletion patterns will frequently result in one or more words being left intact. As a result, *S* has some reliable cues about what must have preceded the intact words and what prob-

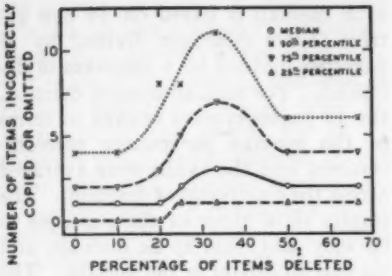


FIG. 6. Number of items incorrectly copied or omitted by the 91 *Ss* in this study. Data for regular and random deletion patterns are combined so that the distribution above each abscissa value has 182 cases, except above the abscissa value 0 where the *N* is 91.

ably follows them. With high amounts of deletion, the regular patterns result in mutilation of every word. For example, the 50% regular deletion pattern takes out every other letter, space, or punctuation mark. This means that every word is affected.

Number of items incorrectly copied.—As noted earlier, *Ss* were instructed to copy what they saw even if they were unable to make any sense of the passage. The reason for this instruction was to insure that every *S* looked at every line of the text. In many instances the very act of copying enabled *S* to see hidden words which he did not see when he first studied the passage. As might have been anticipated, Fig. 6 shows that few *Ss* were able to copy the material presented without error. If the median values in Fig. 6 are converted to percentages of the undeleted text to be copied, the curve is a generally increasing one. This means that percentagewise *Ss* tend to make more copying errors as the text becomes more incomprehensible, that is, as more text is deleted. The data in Fig. 6 combine two sorts of errors—items omitted or items incorrectly copied. Copying errors are essentially identical for random and regular deletion patterns. None of the errors discussed in Fig. 6 are included in any of the foregoing discussion.

Number of items supplied beyond the end of the passage.—Table 1 shows that most *Ss* tried to guess at least a few items beyond the end of the control passages (no text deleted). The number of such guesses decreases rapidly as more and more text is deleted. When 50% or more of the text is deleted, only an occasional *S* was willing to guess at additional ma-

TABLE 1

NUMBER OF ITEMS ATTEMPTED AND NUMBER OF ITEMS CORRECTLY SUPPLIED BEYOND THE END OF THE PASSAGE

(For the control passages $N = 91$; for all other columns $N = 182$)

Measure		Percentage of Items Deleted						
		0 (Control)	10	20	25	33	50	67
No. items attempted	P_{90}	46	47	47	29	8	0	0
	Q_3	36	23	14	8	2	0	0
	Median	9	6	4	0	0	0	0
	Q_1	2	0	0	0	0	0	0
No. items correct	Highest value	16	16	12	10	10	3	2
	Q_3	3	3	3	1	0	0	0
	Median	2	1	1	0	0	0	0
	Q_1	0	0	0	0	0	0	0

terial. This is not a surprising finding; what is striking is that so few of the guesses were correct. For the control passages, the median number of items correctly added to the end of the passage is only two. The maximum number added by any *S* is 16. When 25% or more of the text is deleted, the median number of items correctly added to the end of the text drops to zero.

There were also some large differences between passages in the number of items which could be added to them correctly. No one, for example, was able to guess any items correctly for Passages 8 and 10. On the other hand, 14 *S*s were able to add 10 correct items to the end of Passage 5; 3 *S*s applied 12 correct items; and 1 *S*, 16. Many guesses were also for exactly the same number of items.

These findings can be explained not so much by the relative difficulty of the various passages as by the chance factor of where a passage ends. Passage 8, for example, ends with the end of a sentence. The next sentence starts, "This is a good thing . . ." and no one was able to guess it. Passage 5, on the other hand, ends with ". . . Ted Calaman, of Ph" and many *S*s were able to see that this should be ". . . of Philadelphia, and . . ." Similarly, Passage 7 ends with ". . . and his shoulders seemed, perhaps, a trifle broader than they really . . ." Many *S*s had no difficulty in guessing that the next word is "were." Chance factors like these account for the "all-or-none" nature of the data.

Differences between passages.—There were some large, and fairly consistent, differences between passages in terms of their difficulty. To quantify these differences, tabulations were prepared of the median percentage of items correctly restored for each passage and for each amount of deletion separately. This means that each median is based on 14 raw entries (1,092 messages divided by 13 passages divided by 6 amounts of deletion). For each amount of deletion, the 13 passages were ranked in terms of the median percentage correctly restored and the ranks were averaged across the 6 amounts of deletion. The results show three of the passages to be very easy, four to be difficult, and the rest to cluster in the middle. The easiest passage was Passage 2 (average rank 1.8); Passages 6 and 8 were next easiest (average ranks of 3.7 and 3.8, respectively). Passage 13 was the most difficult (average rank of 10.5), with Passages 10, 3, and 9 rating very high in difficulty (average ranks of 9.2, 8.9, and 8.8, respectively). The differences among these passages are significant at beyond the .01 level as determined by a method of analysis of variance for ranked data.

The extent of these differences is illustrated in Fig. 7 which shows, for

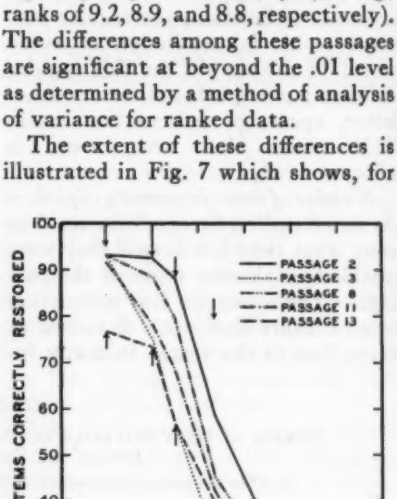


FIG. 7. Percentages of items correctly restored for 5 of the 13 passages in this study. The tips of the arrows show the ranges for all 13 passages.

five selected passages, the median percentage of items correctly restored. Shown here are two of the most difficult, two of the easiest, and one of medium difficulty (Passage 11). To have plotted the results for all 13 passages would have cluttered up the chart without adding any further useful information. A comparison of the amount of variation between passages with that between individuals (Fig. 3) shows that the former is sizeable. The range of variation shown in Fig. 7 is generally greater than the interquartile range for *Ss*.

The rank orders of the 13 passages for every amount of deletion were correlated with the rank orders of the 13 passages for every other amount of deletion (there are a total of 15 such correlations), and the average rank-order correlation was then computed. This average value turns out to be .31. Part of the reason for this low intercorrelation arises from the small spread of values for the 67% deletion (see Fig. 7). It was so difficult that no one could do very much with it. Although not so apparent from Fig. 7, exactly the same situation held for the 10% deletions. The median values for most of these passages were concentrated in the narrow range of 90% to 95%. This amount of deletion was so easy that it did not discriminate among the various passages. The average rank-order correlation computed for the four middle amounts of deletion (that is, 20, 25, 33, and 50%) is higher (.51) indicating that there is a sizeable amount of consistency in the relative rankings of the 13 passages.

Individual differences.—As noted earlier, there were wide differences between *Ss* in their ability to reconstruct these passages. One aim of this study was to discover how well the ability to reconstruct abbreviated messages correlates with general mental ability and with general facility in grammar, spelling, and sentence construction. The measure of general intelligence selected was the Wonderlic test. This test is a highly verbal one, and the correlation between scores on it and scores on the Cooperative English Test, unfortunately, was much higher than anticipated (.75). In an attempt to get a measure which would be somewhat less heavily loaded with verbal factors, I selected 23 of the 50 items on the Wonderlic test which seemed (purely on an intuitive basis) to be less verbal than the remaining 27 items. The score on these 23 items constitutes the "nonverbal Wonderlic score" referred to in Table 2. The score on the remaining 27 items is the "verbal Wonderlic score."

TABLE 2

INTERCORRELATIONS AMONG SCORES

(N = 91 in every case)

Score	(3)	(4)	(5)	(6)
(1) Total Wonderlic test		+.75	+.65	+.38
(2) Nonverbal Wonderlic test	+.63	+.67	+.63	+.37
(3) Verbal Wonderlic test (test 1 minus test 2)		+.71		
(4) Total scaled score on Cooperative English Test			+.68	+.52
(5) Total number of items correctly inserted				
(6) Total number of items added to the end of the passage				

Table 2 shows that the nonverbal score correlates a little lower with the English test, but the difference is so small as to be negligible. If we consider the correlation between the "nonverbal" and "verbal" scores (.63) to be a kind of measure of reliability,

and correct this by the Spearman-Brown formula, the result is an estimated reliability for the Wonderlic of .77. Thus, the Wonderlic and Cooperative English tests correlate with each other about as well as the Wonderlic correlates with itself. It looks as though the two tests are measuring very nearly the same kind of abilities.

We computed two crude measures of each *S*'s ability to reconstruct these abbreviated texts. One was the total number of items inserted correctly into all 12 passages in each person's test booklet. The scores ranged from about 200 to 650, thus indicating a wide range of ability on this task. The second score was the total number of items correctly added beyond the end of the passage summed for the 12 passages in each test booklet. The scores here did not show as great a range. The poorest four *Ss* had scores of 0; the best two had scores of 27.

Table 2 gives the correlations between several of these measures. Of special interest are the high correlations between scores on the two standard tests and the total number of items correctly inserted. In general, scores on the English test correlate a little better with both measures of ability to reconstruct these passages than do scores on the Wonderlic test. This general relationship is maintained in the partial correlations between the English test and number correctly inserted (with scores on the Wonderlic partialed out), and in the correlation between the Wonderlic and number correctly inserted (with scores on the Cooperative English Test partialed out). If we use numerical subscripts with the meanings defined in Table 2, these partial correlations are:

$$r_{45.1} = +.38,$$
$$r_{15.4} = +.26,$$
$$r_{25.4} = +.32.$$

There are two other respects in which general verbal ability shows up in this task. (*a*) Subjects who score high on the paper and pencil tests have a higher ratio of correct items to total items inserted (i.e., the function shown in Fig. 2) than do *Ss* who score low on the tests. These differences, however, appear only for the easy passages. For texts from which a half or two-thirds of the material has been deleted, high-scoring *Ss* are indistinguishable from low-scoring ones. (*b*) Subjects who score high on the paper and pencil tests supplied many more items for passages from which a third or a half of the text had been deleted (this is the function shown in Fig. 1) than did *Ss* who score low on the tests. The differences between *Ss* are not nearly so marked for the easy passages, i.e., those from which less than 25% of the material had been deleted.

Thus the explanation for the correlation between scores on the paper and pencil tests and the number (or percentage) of items correctly restored (i.e., the function shown in Fig. 3) seems to be this: High-scoring *Ss* do better on the easy passages because a higher percentage of the items they supply is correct; for the difficult passages they do better primarily because they supply more items.

The net result of this analysis has been to show that the ability to reconstruct abbreviated English passages is highly correlated with a person's general familiarity with the English language, and that both of these are correlated with general mental ability as measured by a verbal intelligence test. Another way of summing up this analysis is to say that the reconstruction of abbreviated English passages constitutes a kind of intelligence test in itself. From these findings it is perhaps not unreasonable to suggest that persons of high

mental ability can do without some of the redundancy in the English language, but that persons of lesser ability need to have whatever redundancy there is in the language.

Performance of the code experts.— The performance of the code experts was better than that of the median *S* in all of the respects discussed in the section above. However, as noted earlier, the code experts had higher than average scores on both paper and pencil tests. Thus, the question remains, "Do the code experts do better than the median *S* because they are more intelligent than average, or because they have had some specialized training in solving codes?" One way of answering this question is to compare the five code experts with five *S*s who are not code experts, but who match the experts in test scores. Table 3 shows the scores on the Wonderlic and English tests for the code experts and for five other matching *S*s. Although the means for the two groups are almost identical, it was not possible to match *S*s individually.

Figure 8 shows the median, highest, and lowest scores for the code experts and the five matched *S*s. Considering the small numbers involved, the correspondence between the two sets of data is remarkable. Although we cannot make any statements about causes, Fig. 8 shows that code experts do not do any better than *S*s without specialized training in solving codes so long as the latter match the code experts in general verbal ability.

TABLE 3

TEST SCORES FOR THE FIVE CODE EXPERTS AND THE FIVE OTHER MATCHING *S*s

Code Experts		Matching *S*s	
Wonderlic Score	Total English Score	Wonderlic Score	Total English Score
43	71	41	76
39	74	40	75
38	76	36	74
34	79	36	74
29	60	30	60
Means 36.6	72.0	36.6	71.8

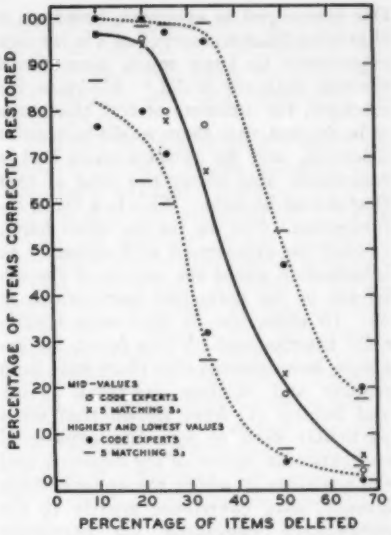


FIG. 8. Percentage of items correctly inserted by the five code experts and by five *S*s matched with the code experts in Wonderlic score and total English score

DISCUSSION

One important finding of this study is that people cannot do a very good job of reconstructing English prose passages when the amount of material deleted is 30% or more. Although there are marked differences between *S*s, the very best performance by *any S* on *any* passage showed only 60% recovery for text from which 50% had been deleted. These findings are in strong contrast to the statement which Shannon has made about this kind of task.

There are many possible reasons why these results differ so markedly from Shannon's. Some of these sources of discrepancy are evident from the findings of this study, e.g., the kind of passage selected for test and the verbal ability of the *S* are important considerations.

One reason not so apparent, however, is that when Shannon served as *S* in his own experiment he knew much more about the task than our *S*s did.[5] He knew, for example, the amounts of text that were to be deleted, that there would be regular deletions, and he perhaps even had a reasonable idea about the kind of text that would be used. This is a lot of information. Our *S*s, on the other hand, entered the experiment with virtually no information about the nature of the deletions or the texts (see instructions to *S*). Of those few *S*s who were specifically interrogated on this point, for example, none guessed that there were both regular and random deletions. Frick and Sumby (1) have shown that for a particular kind of job, the situational contexts (the nature of the audience and the situation in which the speaker finds himself) may contribute greatly to the redundancy computed for particular samples of language. The same should be true of explicit or implicit instructions. In short, if our *S*s had been told, "This is a recipe for apple pie from which a third of the items have been deleted in a random manner," there seems to be little reason to doubt that they would have performed better than they did. The precise contribution of such additional contextual constraints has yet to be evaluated.

There is, of course, the important question of motivation. Shannon worked under extremely high motivation.[6] We think that at least some of our *S*s were highly motivated for this task. Many of them regarded it as a challenge since it has something of the flavor of anagrams, or double crostics. As evidenced from their comments, some of the *S*s put their best efforts into the task. Other *S*s, of course, were easily discouraged despite instructions and inducements. There can be no question that motivation is a major, uncontrolled variable in this study.[7]

This study has been directed toward only one aspect of the problem of redundancy and how it helps in the reconstruction of prose passages. There is little doubt but that the task could have been made much simpler by providing *S* with additional cues. If, for example, blank spaces had been put in wherever something was deleted, the task would have been comparable to double crostics, and there is no doubt that performance should be much better.

There are two other aspects of the

[5] Shannon, C. E. Personal communication.

[6] See footnote 5.

[7] *Note added in proof*: In June 1954 a copy of this manuscript was sent to Dr. Shannon for his criticism. Upon reading it Dr. Shannon felt that the differences between his observations and mine were probably due to the fact that the *S*s in this study were insufficiently motivated and that they had not spent enough time at the task. As a result I prepared at that time two copies each of four abbreviated passages (different from any of those in this study) for his use. Two had 50% deleted in a regular pattern; one had 50% deleted randomly; one had 60% deleted in a regular pattern.

In a letter dated November 17, 1954, Dr. Shannon communicated the results of some additional tests. He and Mrs. Shannon had independently attempted the reconstruction of the four passages. For the two passages from which 50% had been deleted in a regular pattern, they were able to restore from 83% to 93% of the deleted content. For the other two passages neither he nor Mrs. Shannon was able to reconstruct "anything significant." On the basis of these and still other tests, Dr. Shannon is now inclined to believe that the differences between our observations are not so much a matter of motivation as they are the result of the amount of supplementary information *S* has. He and Mrs. Shannon were able to do well on two of the passages (and in a relatively short time) because they looked for and discovered the simple, regular deletion pattern.

Dr. Shannon and I now agree that *S*s can probably reconstruct from about 80% to 100% of deleted text under the following special conditions: (*a*) The amount deleted is 50%; (*b*) The deletions are made by taking out every other space, letter, or punctuation mark (other kinds of 50%, regular deletion patterns are more difficult to reconstruct); (*c*) The *S* is told, or can easily discover, the deletion pattern; (*d*) The *S* has a high amount of verbal intelligence.

The results of the present study are valid under these conditions: (*a*) The *S* is provided with no supplementary information about the amount or kind of deletion; (*b*) The total context of the situation is such that *S* cannot or does not discover the deletion patterns.

findings which are difficult to quantify, but which are so striking that they deserve mention. One is the exceptional capacity some *S*s demonstrated for reconstructing material from extremely meager cues. The second is that the intelligibility of the messages seemed to be seriously affected when a third or more of the text had been deleted. By intelligibility I mean that *S*s obviously misinterpreted the subject matter of the passage. It is difficult to know when, or how frequently, this occurs. The impression one gets is that intelligibility starts to suffer seriously when a third or more of the text has been deleted.

Perhaps the best way of documenting these points is to cite some actual results. The performance of one *S* on Passage 11 illustrates both points well. The passage had 50% deleted in a regular pattern. The first two lines presented to the *S* read "I stertclyntipsil odvlp tesaitc." From this extremely sketchy text, she was able to reconstruct the first nine words perfectly, but thereafter it was apparent that she lost the context. Her reconstruction began: "It is theoretically not impossible to develop the statistics for testing men in . . ." (compare this with the original).

Another *S*, working with Passage 9 from which 50% of the text had been randomly deleted, reconstructed the phrase "Pare, core, and cut . . ." as "Prices, wages, dollars and cents . . ." This set for content relating to an economic subject was evident throughout the rest of the passage. One of the most striking examples of the loss of intelligibility occurred with Passage 6, from which two-thirds of the text had been deleted in a regular pattern. One *S* reconstructed the first two lines as: "Ruins of Greece recently discovered tell that the ancient Hellonics are . . ." These examples are perhaps sufficient to support the qualitative finding that *S*s oftentimes completely lost the content of material from which 50% or more of the text had been deleted. This occurs so often that we find it difficult to believe that Shannon's report is typical of what one might expect the average person to do. It would be interesting to try to get some objective evidence of this loss of intelligibility (perhaps by questions about the content of the passage), but this constitutes another study.

The final point of interest concerns the relative difficulty of the passages in a task such as this. Material which is easy to read (as measured by a Flesch count, for example) is not necessarily easy to reconstruct under the conditions of this experiment. Passage 2, for example, was the easiest passage to reconstruct and Passage 6 was next easiest. Yet both of these are clearly more difficult reading than Passage 7. Passages 2 and 6 contain many long words of moderate difficulty. Passage 7 was relatively difficult to reconstruct primarily because it contained too many short words which lost their identity as soon as some of the text had been deleted. Passage 9 is relatively easy reading, but it was difficult to reconstruct because so much redundancy had already been taken out of it. There is no "padding" in Passage 9.

Summary and Conclusions

This experiment was aimed at discovering how well a heterogeneous sample of *S*s can reconstruct English text when various amounts of it have been deleted. Thirteen prose passages, differing widely in content and style, were selected for test. Deletions were made in both regular and random patterns and in six amounts varying from 10% to 67%. Ninety-one *S*s, varying widely in general intelligence and familiarity with grammar, were used in the study. Five of the *S*s were naval code experts. The *S*s were essentially instructed to reconstruct the original passage from the material presented. The results show that:

1. The typical *S* supplies about the correct number of items when the amount deleted is 25% or less. For greater amounts of deletion too little material is supplied.

2. A relatively small amount of the material supplied is correct. When 10% of the text has been deleted, only about 80% of the items supplied are correct. The percentage decreases markedly for greater amounts of deletion.

3. The percentage of material which can be correctly restored drops off very rapidly as more and more text is deleted. In general, the typical

S cannot recover very much material. When 25% of the text has been deleted, only about 70% of it is recovered. When 50% of the text has been deleted, only about 10% can be recovered. There are, however, wide individual differences in this respect.

4. When the amount of deletion is 25% or less it is more difficult to reconstruct randomly deleted passages than those which were deleted in a regular pattern. When the amount of deleted material exceeds 33%, randomly deleted passages are easier to reconstruct.

5. Most *S*s were not able to add very much beyond the content supplied them.

6. There are sizeable differences between passages in the ease with which they can be reconstructed. In general, passages which are easy to reconstruct are not necessarily those which are easiest to read according to conventional readability criteria.

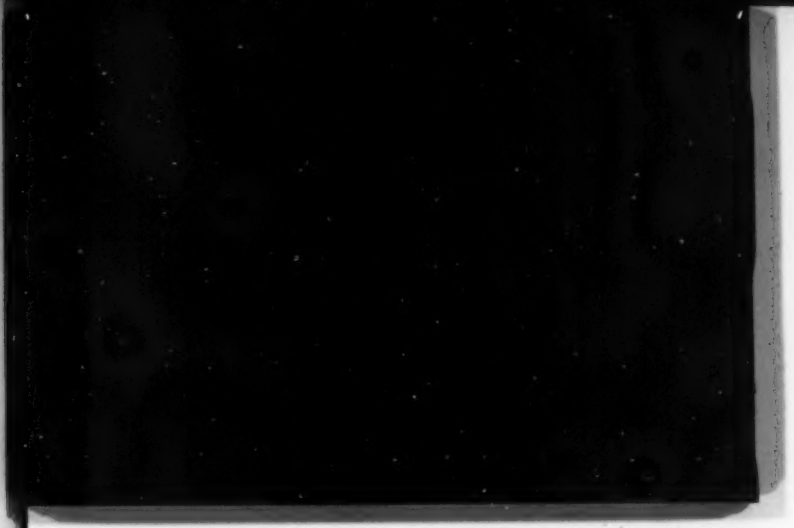
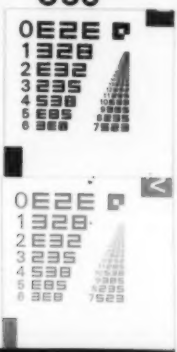
7. As a group, the code experts did considerably better than the median *S*. Even they, however, were not able to reconstruct very much material when the amount of deletion was 50% or greater. The code experts do not perform better than *S*s who match them in general verbal and mental ability.

8. The ability to reconstruct passages in this experiment correlates highly with scores on the Wonderlic Personnel Test and the Cooperative English Test (*r*'s range from .63 to .68).

References

1. Frick, F. C., & Sumby, W. H. Control tower language. *J. acoust. Soc. Amer.*, 1952, **24**, 595–596.
2. Shannon, C. E., & Weaver, W. *The mathematical theory of communication.* Urbana, Ill.: Univ. Illinois Press, 1949.

(Received for early publication June 7, 1954)



C36